

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

STIC-Biotech/ChemLib

100 291

From: Nickol, Gary
Sent: Friday, August 01, 2003 11:24 AM
To: STIC-Biotech/ChemLib
Subject: 09/988975

RECEIVED

AUG - 1 2003

Please search and interference search the following:

(STIC)

SEQ ID NO:1

Thanks,

Gary Nickol
CM1, AU:1642
Room 8D17, Mailbox 8E12
(703) 305-7143

Point of Contact
P. Sheppard
Telephone number: (703) 308-4499

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: 8/4/03
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2003, 18:10:54 ; Search time 19 seconds
(without alignments)
630.209 Million cell updates/sec

Title: US-09-988-975A-1
Perfect score: 1543

Sequence: 1 MKLNTSAGNVDIYKKLYHSD.....VYGNVWVFYDWIRQBADG 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/6C_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/6D_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 1541 | 99.9 | 283 | 3 | US-08-807-151-1 |
| 2 | 1541 | 99.9 | 283 | 3 | US-09-478-957-1 |
| 3 | 1540 | 99.8 | 492 | 3 | US-09-342-749-2 |
| 4 | 1540 | 99.8 | 492 | 3 | US-09-691-840-2 |
| 5 | 857 | 55.5 | 159 | 3 | US-09-518-046-24 |
| 6 | 696 | 45.1 | 454 | 3 | US-09-518-046-2 |
| 7 | 588.5 | 38.1 | 455 | 3 | US-09-261-416-2 |
| 8 | 558.5 | 36.2 | 423 | 3 | US-09-656-002-2 |
| 9 | 558.5 | 36.2 | 435 | 3 | US-09-008-271A-6 |
| 10 | 544.5 | 35.3 | 248 | 3 | US-08-944-483-63 |
| 11 | 541 | 35.1 | 738 | 2 | US-08-681-151-3 |
| 12 | 536.5 | 34.8 | 798 | 5 | US-08-200-900A-2 |
| 13 | 523.5 | 33.9 | 376 | 5 | PCT-US94-00616-2 |
| 14 | 523.5 | 33.9 | 376 | 5 | US-09-820-002-2 |
| 15 | 520.5 | 33.7 | 417 | 4 | US-09-820-002-4 |
| 16 | 519.5 | 33.7 | 256 | 2 | US-09-027-337-3 |
| 17 | 519.5 | 33.7 | 256 | 4 | US-09-644-600-3 |
| 18 | 518.5 | 33.6 | 418 | 1 | US-08-508-448C-25 |
| 19 | 518.5 | 33.6 | 418 | 4 | US-09-370-838-62 |
| 20 | 518.5 | 33.6 | 418 | 4 | US-09-370-838-82 |
| 21 | 518.5 | 33.6 | 418 | 4 | US-09-370-838-83 |
| 22 | 515 | 33.4 | 238 | 3 | US-08-944-483-64 |
| 23 | 514.5 | 33.3 | 235 | 3 | US-08-944-483-67 |
| 24 | 512.5 | 33.2 | 235 | 3 | US-08-807-151-3 |
| 25 | 512.5 | 33.2 | 235 | 4 | US-09-478-957-3 |
| 26 | 512.5 | 33.2 | 855 | 2 | US-09-027-337-2 |
| 27 | 512.5 | 33.2 | 855 | 4 | US-09-644-600-2 |

| | | | | | | |
|----|-------|------|-----|---|-------------------|--------------------|
| 28 | 505.5 | 32.8 | 812 | 1 | US-08-248-629A-1 | Sequence 1, Appl1 |
| 29 | 505.5 | 32.8 | 812 | 1 | US-08-451-932-1 | Sequence 1, Appl1 |
| 30 | 505.5 | 32.8 | 812 | 1 | US-08-452-260-1 | Sequence 1, Appl1 |
| 31 | 505.5 | 32.8 | 812 | 1 | US-08-326-785-1 | Sequence 1, Appl1 |
| 32 | 505.5 | 32.8 | 812 | 2 | US-08-612-788-1 | Sequence 1, Appl1 |
| 33 | 505.5 | 32.8 | 812 | 2 | US-08-605-598B-1 | Sequence 1, Appl1 |
| 34 | 505.5 | 32.8 | 812 | 2 | US-08-429-743-1 | Sequence 1, Appl1 |
| 35 | 505.5 | 32.8 | 812 | 2 | US-08-866-735-1 | Sequence 1, Appl1 |
| 36 | 505.5 | 32.8 | 812 | 3 | US-09-066-028-1 | Sequence 1, Appl1 |
| 37 | 505.5 | 32.8 | 812 | 4 | US-09-192-012-3 | Sequence 3, Appl1 |
| 38 | 505.5 | 32.8 | 812 | 4 | US-09-335-325-1 | Sequence 1, Appl1 |
| 39 | 505.5 | 32.8 | 812 | 4 | US-08-991-761A-12 | Sequence 12, Appl1 |
| 40 | 505.5 | 32.8 | 812 | 5 | PCT-US95-05107-1 | Sequence 1, Appl1 |
| 41 | 504.5 | 32.7 | 232 | 1 | US-08-508-448C-19 | Sequence 19, Appl1 |
| 42 | 502.5 | 32.6 | 235 | 3 | US-08-944-483-65 | Sequence 65, Appl1 |
| 43 | 501.5 | 32.5 | 916 | 2 | US-09-000-846-2 | Sequence 2, Appl1 |
| 44 | 498.5 | 32.3 | 902 | 4 | US-09-644-600-10 | Sequence 10, Appl1 |
| 45 | 491.5 | 31.9 | 285 | 4 | US-09-023-942A-26 | Sequence 26, Appl1 |

ALIGNMENTS

RESULT 1
US-08-807-151-1
Sequence 1, Application US/08807151
Patent No. 6043033
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
TITLE OF INVENTION: PROTEASE
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESSES:
ADDRESSER: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/807,151
FILING DATE: Filed Herewith
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0227 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 283 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SCORNOT01
CLONE: 556016
US-08-807-151-1

Query Match 99.9%; Score 1541; DB 3; Length 283;
Best Local Similarity 100.0%; Pred. No. 8.5e-161;
Matches 283; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLNTSAGNVDIYKLLYHSDACSSKAVVSLRCLACGVNLNSRSRQRIYVGBSALPGAMPW 60
DB 1 MKLNTSAGNVDIYKLLYHSDACSSKAVVSLRCLACGVNLNSRSRQRIYVGBSALPGAMPW 60
QY 61 QVSLHVQVNVHVCSSIIITPEWIVTAHCVKELNPNMHTAFAGILROSFMFYGAGYOYE 120
DB 61 QVSLHVQVNVHVCSSIIITPEWIVTAHCVKELNPNMHTAFAGILROSFMFYGAGYOYE 120
QY 121 KVISHPNVDSTKKNDDIALMKLOKPLTFNDLVKPVCLPMPGMLOPEOLCWSIGMGATEE 180
DB 121 KVISHPNVDSTKKNDDIALMKLOKPLTFNDLVKPVCLPMPGMLOPEOLCWSIGMGATEE 180
QY 181 KGKTSBVNAAKVLLIETQRCNSRYVYDNLITPAMI CAGFLQGNVDS CGDSGGXLYTSK 240
DB 181 KGKTSBVNAAKVLLIETQRCNSRYVYDNLITPAMI CAGFLQGNVDS CGDSGGXLYTSK 240
QY 241 NNIMWLIGDTSWSSGCAKAYRPGVYGWVFTDWTYRQBRADG 283
DB 241 NNIMWLIGDTSWSSGCAKAYRPGVYGWVFTDWTYRQBRADG 283

RESULT 2
US-09-478-957-1
; Sequence 1, Application US/09478957
; Patent No. 6350448
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
; TITLE OF INVENTION: PROTEASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/478,957
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/807,151
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0227 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-845-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 283 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SCORNO701
; CLONE: 556016
; US-09-478-957-1

Query Match 99.9%; Score 1541; DB 4; Length 283;
Best Local Similarity 100.0%; Pred. No. 8.5e-161;
Matches 283; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLNTSAGNVDIYKLLYHSDACSSKAVVSLRCLACGVNLNSRSRQRIYVGBSALPGAMPW 60
DB 1 MKLNTSAGNVDIYKLLYHSDACSSKAVVSLRCLACGVNLNSRSRQRIYVGBSALPGAMPW 60
QY 61 QVSLHVQVNVHVCSSIIITPEWIVTAHCVKELNPNMHTAFAGILROSFMFYGAGYOYE 120
DB 61 QVSLHVQVNVHVCSSIIITPEWIVTAHCVKELNPNMHTAFAGILROSFMFYGAGYOYE 120
QY 121 KVISHPNVDSTKKNDDIALMKLOKPLTFNDLVKPVCLPMPGMLOPEOLCWSIGMGATEE 180
DB 121 KVISHPNVDSTKKNDDIALMKLOKPLTFNDLVKPVCLPMPGMLOPEOLCWSIGMGATEE 180
QY 181 KGKTSBVNAAKVLLIETQRCNSRYVYDNLITPAMI CAGFLQGNVDS CGDSGGXLYTSK 240
DB 181 KGKTSBVNAAKVLLIETQRCNSRYVYDNLITPAMI CAGFLQGNVDS CGDSGGXLYTSK 240
QY 241 NNIMWLIGDTSWSSGCAKAYRPGVYGWVFTDWTYRQBRADG 283
DB 241 NNIMWLIGDTSWSSGCAKAYRPGVYGWVFTDWTYRQBRADG 283

RESULT 3
US-09-342-749-2
; Sequence 2, Application US/09342749
; Patent No. 6166194
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-P.
; APPLICANT: Wyriad Genetics, Inc.
; TITLE OF INVENTION: TMPRSS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/342,749
; CURRENT FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: US 60/091,044
; EARLIER FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-342-749-2

Query Match 99.8%; Score 1540; DB 3; Length 492;
Best Local Similarity 99.6%; Pred. No. 2.5e-160;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKLNTSAGNVDIYKLLYHSDACSSKAVVSLRCLACGVNLNSRSRQRIYVGBSALPGAMPW 60
DB 210 MKLNTSAGNVDIYKLLYHSDACSSKAVVSLRCLACGVNLNSRSRQRIYVGBSALPGAMPW 269
QY 61 QVSLHVQVNVHVCSSIIITPEWIVTAHCVKELNPNMHTAFAGILROSFMFYGAGYOYE 120
DB 270 QVSLHVQVNVHVCSSIIITPEWIVTAHCVKELNPNMHTAFAGILROSFMFYGAGYOYE 329
QY 121 KVISHPNVDSTKKNDDIALMKLOKPLTFNDLVKPVCLPMPGMLOPEOLCWSIGMGATEE 180
DB 330 KVISHPNVDSTKKNDDIALMKLOKPLTFNDLVKPVCLPMPGMLOPEOLCWSIGMGATEE 389
QY 181 KGKTSBVNAAKVLLIETQRCNSRYVYDNLITPAMI CAGFLQGNVDS CGDSGGXLYTSK 240
DB 390 KGKTSBVNAAKVLLIETQRCNSRYVYDNLITPAMI CAGFLQGNVDS CGDSGGXLYTSK 449
QY 241 NNIMWLIGDTSWSSGCAKAYRPGVYGWVFTDWTYRQBRADG 283
DB 450 NNIMWLIGDTSWSSGCAKAYRPGVYGWVFTDWTYRQBRADG 492

RESULT 4
US-09-691-840-2
; Sequence 2, Application US/09691840
; Patent No. 6444419

GENERAL INFORMATION:
APPLICANT: Wong, Alexander K.C.
APPLICANT: Tavagian, Sean V.
APPLICANT: Teng, David H.-F.
APPLICANT: Myriad Genetics, Inc.
TITLE OF INVENTION: TMPRSS2 is a Tumor Suppressor
FILE REFERENCE: 2318-202
CURRENT APPLICATION NUMBER: US/09/691,840
CURRENT FILING DATE: 2000-10-18
PRIOR APPLICATION NUMBER: US/09/342,749
PRIOR FILING DATE: 1999-06-29
PRIOR APPLICATION NUMBER: US 60/091,044
PRIOR FILING DATE: 1998-06-29
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 492
TYPE: PRT
ORGANISM: Homo sapiens
US-09-691-840-2

Query Match 99.8%; Score 1540; DB 4; Length 492;
Best Local Similarity 99.6%; Pred. No. 2, 5e-160;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MLNTSAGVNDLYKCLYHS DACSSKAVSLRCIACGVNLSRSRSTVGGESALPGAMPW 60
DB 210 MLNTSAGVNDLYKCLYHS DACSSKAVSLRCIACGVNLSRSRSTVGGESALPGAMPW 269
QY 61 QVSLHVQNVHVCSSGSIITPEMTVTAHCEKPLNNPMTAFAGILRSFMYGAGYOV 120
DB 270 QVSLHVQNVHVCSSGSIITPEMTVTAHCEKPLNNPMTAFAGILRSFMYGAGYOV 329
QY 121 KVISHPNYSKTKNDIALMKLOKPLTFNDLVKPVCLPFGMWLOPEOLCWSGMAATE 180
DB 330 KVISHPNYSKTKNDIALMKLOKPLTFNDLVKPVCLPFGMWLOPEOLCWSGMAATE 389
QY 181 KKTSEVTLNAAYLLIETORCNSRYVDLITPAMICAGFLQGNVSCGDSGSKLVTSK 240
DB 390 KKTSEVTLNAAYLLIETORCNSRYVDLITPAMICAGFLQGNVSCGDSGSKLVTSK 449
QY 241 NNIMWLIGDTSWGSCKAKYRPGVGNVVFDTMIRORADG 283
DB 450 NNIMWLIGDTSWGSCKAKYRPGVGNVVFDTMIRORADG 492

RESULT 5
US-09-518-046-24
Sequence 24, Application US/09518046
Patent No. 6294663
GENERAL INFORMATION:
APPLICANT: O'Brien, Timothy J.
APPLICANT: Underwood, Lowell J.
TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
FILE REFERENCE: D6192CIP
CURRENT APPLICATION NUMBER: US/09/518,046
CURRENT FILING DATE: 2000-03-02
EARLIER APPLICATION NUMBER: 09/261,416
EARLIER FILING DATE: 1999-03-03
NUMBER OF SEQ ID NOS: 153
SEQ ID NO 24
LENGTH: 159
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: DOMAIN
OTHER INFORMATION: protease domain of TMPRSS2 (TMPRSS2)
US-09-518-046-24

Query Match 55.5%; Score 857; DB 3; Length 159;
Best Local Similarity 98.7%; Pred. No. 4, 2e-86;
Matches 157; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 80 EMTVTAHCEKPLNNPMTAFAGILRSFMYGAGYOVKVISHPNYSKTKNDIAL 139
DB 1 EMTVTAHCEKPLNNPMTAFAGILRSFMYGAGYOVKVISHPNYSKTKNDIAL 60
QY 140 MKLOKPLTFNDLVKPVCLPFGMWLOPEOLCWSGMAATEBKTSSEVTLNAAYLLIETQ 199
DB 61 MKLOKPLTFNDLVKPVCLPFGMWLOPEOLCWSGMAATEBKTSSEVTLNAAYLLIETQ 120
QY 200 KCKSRYYVDNLITPAMICAGFLQGNVSCGDSGSKLVLT 238
DB 121 KCKSRYYVDNLITPAMICAGFLQGNVSCGDSGSKLVLT 159

RESULT 6
US-09-518-046-2
Sequence 2, Application US/09518046
Patent No. 6294663
GENERAL INFORMATION:
APPLICANT: O'Brien, Timothy J.
APPLICANT: Underwood, Lowell J.
TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
FILE REFERENCE: D6192CIP
CURRENT APPLICATION NUMBER: US/09/518,046
CURRENT FILING DATE: 2000-03-02
EARLIER APPLICATION NUMBER: 09/261,416
EARLIER FILING DATE: 1999-03-03
NUMBER OF SEQ ID NOS: 153
SEQ ID NO 2
LENGTH: 454
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: complete amino acid sequence of TADG-12
US-09-518-046-2

Query Match 45.1%; Score 696; DB 3; Length 454;
Best Local Similarity 50.4%; Pred. No. 9e-68;
Matches 137; Conservative 37; Mismatches 92; Indels 6; Gaps 4;

QY 12 IYKLYHS DACSSKAVSLRCIACGVNLSRSRSTVGGESALPGAMPQVSLHVQNVH 71
DB 185 LHSYVYREGCASGHVVTLLQCTACGHRGYS--SRIVGAGMSLSQMPQASLQFGYHL 242
QY 72 CGGSITPEMTVTAHCEKPLNNPMTAFAGILRSFMYGA-GYOVKVISHPNYS 130
DB 243 CGGSITPEMTVTAHCEKPLNNPMTAFAGILRSFMYGA-GYOVKVISHPNYS 239
QY 131 KTKNDIALMKLOKPLTFNDLVKPVCLPFGMWLOPEOLCWSGMAATEBKTSSEVTLNA 190
DB 300 KTLGNDIALMKLAGPLTFNBMLQPVCLPFSERNFPDGRKCTWSGMAATEBDGDSAPVLAN 359
QY 191 AYVLLIETORCNSRYVDLITPAMICAGFLQGNVSCGDSGSKLVTSKNNIMWLIGDT 250
DB 360 AAVPLISNKICNHRDYGIIISPSMLCAGYLTCGVDS CGDSGSKLVTSKNNIMWLIGDT 419
QY 251 SWSGCAKAYRPGVGNVVFDTMIRORAD 282
DB 420 SFGICARVKNKPGVYTRVTSFLDMIRORAD 451

RESULT 7
US-09-261-416-2
Sequence 2, Application US/09261416A
Patent No. 6291663
GENERAL INFORMATION:
APPLICANT: O'Brien, Timothy J.
APPLICANT: Underwood, Lowell J.
TITLE OF INVENTION: TADG-12: A No. 6291663el Transmembrane Serine Protease
FILE REFERENCE: D6192

CURRENT APPLICATION NUMBER: US/09/261,416A
CURRENT FILING DATE: 1999-03-03
NUMBER OF SEQ ID NOS: 14
SEQ ID NO 2
LENGTH: 455
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Amino acid sequence of TADG-12 encoded by nucleotides
Patent No. 6291663
US-09-261-416-2

Query Match 38.1%; Score 588.5; DB 3; Length 455;
Best Local Similarity 46.2%; Pred. No. 5,7e-56;
Matches 127; Conservative 35; Mismatches 102; Indels 11; Gaps 7;

QY 12 IYKKLYHSDACSSKAVVSLRCLACGVNLSSRQSRIVGSGALPGAMPQVSLHVNQVH 71
DB 185 LHSYVREGCAGSHVYITQCTACGRGYS--SRVGMMSLSSQPMQASLQFGGTH 242
QY 72 CGGSIITPEWITAAHCEKPLNPMWMTAFAGILRQSMFYGA-GYQEKVISHPNYDS 130
DB 243 CGGSIITPEWITAAHCV-YDLYLPSKWTIYQGLV--SLDNPAPSHLVEKIVYHSKYP 299
QY 131 KTKNDIALMKQKPLTENDLVKVCLEPNPMMLOEOLCWSGQATEBEKTSYLA 190
DB 300 KRLGNDIALMKLAGPLTFEMMIQPVCLPNSSENPDPGKVCMTSGMGTEDGSDASPLNH 359
QY 191 AKVLIENQ--RCNSRYVDNLTTPAMICAGFLQG--NVDSGQDSGGLVTSKNNTWLI 247
DB 360 AAVPLISNDLOPQGRVHML--PLMALRLGLPDGRKMSQGDSDGGLVCOERRMLKV 417
QY 248 GDTWSGGCAKAYRPGVYGVVWFTDWYRQMRAD 282
DB 418 GATSPGIGADVKNKGVYTRVTSFLMWHQMRD 452

RESULT 8
US-09-656-002-2
Sequence 2, Application US/09656002
Patent No. 6455668
GENERAL INFORMATION:
APPLICANT: Mack, David
APPLICANT: Wilson, Kurt
TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
FILE REFERENCE: A-69108/DOB/JUD/AMS
CURRENT APPLICATION NUMBER: US/09/656,002
CURRENT FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: US 09/525,993
PRIOR FILING DATE: 2000-03-15
PRIOR APPLICATION NUMBER: US 09/493,444
PRIOR FILING DATE: 2000-01-28
PRIOR APPLICATION NUMBER: PCT/US 00/07044
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 423
TYPE: PRT
ORGANISM: Homo sapiens
US-09-656-002-2

Query Match 36.2%; Score 558.5; DB 4; Length 423;
Best Local Similarity 43.9%; Pred. No. 1e-52;
Matches 118; Conservative 41; Mismatches 93; Indels 17; Gaps 8;

QY 19 SDACSSKAVVSLRCLACGVNLSSRQSRIVGSGALPGAMPQVSLHVNQVHVGGSIT 78
DB 166 SGPCLSGSLVSLHCLACGKSL---KTPRYVGGEEASVDSMPQVSIQYKQHVCGGSIID 222

QY 79 PSHVITAAHCEKPLNPMWMTAFAGILR-QSFMFYGAGYQEKVI---SHRYDSKTRN 134
DB 223 PSHVITAAHCEKPLNPMWMTAFAGILR-QSFMFYGAGYQEKVI---SHRYDSKTRN 134
QY 135 NDIALMKQKPLTENDLVKVCLEPNPMMLOEOLCWSGQATEBEKTSYLA 193
DB 275 NDIALMKQKPLTENDLVKVCLEPNPMMLOEOLCWSGQATEBEKTSYLA 193
QY 194 LLIETQRNSRYVDNLTTPAMICAGFLQGNVDSGQDSGGLVTSKNNTWLI 247
DB 335 QVLDSTRCNADAYQGEVTEKMKCAGIPRGGVDTCCGDSGGLVTSKNNTWLI 247
QY 254 SGCARAYRPGVYGVVWFTDWYRQMRAD 282
DB 394 YGGGSPSTGGVYTRVTSFLMWHQMRD 452

RESULT 9
US-09-008-271A-6
Sequence 6, Application US/09008271A
Patent No. 6203979
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
Hillman, Jennifer L.
Yue, Henry
Guegler, Karl J.
Corley, Neil C.
Tang, Tom Y.
Shah, Purvi
TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSER: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/008,271A
FILING DATE: 16-Jan-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mohan-Peterson, Sheela
REGISTRATION NUMBER: 41,201
REFERENCE/DOCKET NUMBER: PF-0458 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: COLN0013
CLONING: 1337018
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-008-271A-6

Query Match 36.2%; Score 558.5; DB 3; Length 435;
Best Local Similarity 43.9%; Pred. No. 1e-52;
Matches 118; Conservative 41; Mismatches 93; Indels 17; Gaps 8;

QY 19 SDACSSKAVVSLRCLACGVNLSSRQSRIVGSGALPGAMPQVSLHVNQVHVGGSIT 78

Db 178 SGPCLSGSLVSLHCLAGSBL---KTPRVVGGEEASVDSWPMQVSIQYKQKHVCGGSILD 234
Qy 79 PEMVITAAICVEKPLNNPMTAFAGILR--OSPMFYGAGYQVEKVI---SHPNVDSKTKN 134
Db 235 PMVLTAAICFRGH-IDVPMKTRAGSDKLSF---PSLAVAKIITIERPMI---PKD 286
Qy 135 NDIALMKLQPLTFSGVTRPCLPFDEBELTPATPLMIIGMGFTKONGKMSDILLQASV 346
Db 287 NDIALMKLQPLTFSGVTRPCLPFDEBELTPATPLMIIGMGFTKONGKMSDILLQASV 346
Qy 194 LLETRCRKSRYYINDLITPAMICAGFLQGNVDSGGDSGKLVTSKNNIWMILGDTSWG 253
Db 347 QYIDSTRCNADDAVYQSEVTEKMKACGIPGCGVDTCQDGSAGPLMYQSDQ--MHVGVIVSWG 405
Qy 254 SGCARAYRPGVGNVWFTDMYRQMRAD 282
Db 406 YGCGGPRSTGVYTKVSAVLMVYVWKA 434

RESULT 10

US-08-944-483-63
; Sequence 63, Application US/08944483
; Patent No. 6232456
; GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: KLAS, MICHAEL R.
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STEWART, KENT D.
; APPLICANT: STROUPE, STEVEN D.
; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,483
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6183.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELETYPE:
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6232456e
; US-08-944-483-63

Query Match 35.3%; Score 544.5; DB 3; Length 248;
Best Local Similarity 42.8%; Pred. No. 1.6e-51;
Matches 104; Conservative 46; Mismatches 84; Indels 9; Gaps 6;

Qy 47 IUGESALPAPMPQVSLHVO---NHVCGSITTPMIVTAACV-KEPLNNPMTAF 102
Db 1 IUGVNSWGEWPMQVSLQVLTQNRHLCGSLIGHQWVTAACFDGLPQDWR--LY 58
Qy 103 AGILRQSPMYGAGY--QVEKVI SHPNVDSKTKNDILAMQLQKELTRNDLVKPCLPKPG 161
Db 59 SGLINLSIDTKDTPPSQIKKIIHQNKVSEGNHDLITLQAPLANTFEPKPICLPKG 118
Qy 162 MMLQPEOLCWSGNGATEEKGKTSVYNAKAVLLIETORCNSRYVYNDLITPAMICAGFL 221
Db 119 DSTIYTNCAVYTGKPSKEKEIIONILQKNIPLVNTECKRI--QDYKTIQRMVCAGYK 177
Qy 222 QGNVDSGGDSGKLVTSKNNIWMILGDTSWGSCARAYRPGVGNVWFTDMYRQMR- 280
Db 178 EGKDKACKDSDGGLVCKHGMRLVGIITSWGECARREQGVYTTVAIVMDWILEKTS 237
Qy 281 ADG 283
Db 238 SDG 240

RESULT 11

US-08-681-151-3
; Sequence 3, Application US/08681151
; Patent No. 5869637
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Bandman, Olga
; APPLICANT: Braxton, Scott Michael
; APPLICANT: Goll, Surya
; TITLE OF INVENTION: A NOVEL HUMAN KALLIKREIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/681,151
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0074US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELETYPE:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 638 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: 205011

US-08-681-151-3

Query Match 35.1%; Score 541; DB 2; Length 638;
Best Local Similarity 38.2%; Pred. No. 1.5e-50;
Matches 109; Conservative 54; Mismatches 112; Indels 10; Gaps 5;

QY 1 MKLNTSAGNVDIYKLYHSDACSSKAVVSLRCIACGVNLNSSROS-RIVGESALPGAW 60
DB 349 URLSTDSPTRTTYAOGSSGYSLRCKYVSSDCTTKIN---ATVGTGTSJGBEWP 404
QY 61 QVSLHVQNV---HVCGSITTPETVITAAHCYKELNPMHTAFAGILROSMFYGAGY 117
DB 405 QVSLQVTKVSNHMGCGSITIGRWLITAHCFD-GIPYDVIRIYGILNLSITNKTF 463
QY 118 -QVEKVISHPNYDSKTKNDIALMKLQKPLTFNDLVKVCPLNPGMLOPEOLCWSMG 176
DB 464 SSIKELIHOKYKMGSEGYDIALIKLQPLNTYTERFOKPLCPSKADTNTIYTNCWATG 523
QY 177 ATEBKTSSEVLNAKVLLEFORCNSRYVDNLITPMICAGFLQGNVDSQGSQGL 236
DB 524 YTKERGETONILQKATIFLVNPECOKY-RDYVITKMICAGYKGGIDACKDSGGL 582
QY 237 VTSKNIMWLIGDTSMGSCAKAYRPGYGVNMTFTWYQWRA 281
DB 583 VCKSRKQVLGVTSMGSCAKRQPGYVTKYAEYIDLEKIG 627

RESULT 12
US-08-200-900A-2

Sequence 2, Application US/08200900A

Patent No. 5665566

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESS:

ADDRESS: Genetics Institute, Inc. - Legal Affairs

STREET: 87 Cambridge Park Drive

CITY: Cambridge

STATE: MA

COUNTRY: USA

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/200,900A

FILING DATE: 23-FEB-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Meiner, Maureen C.

REGISTRATION NUMBER: 31,544

TELEPHONE: (617) 876-1170 X8574

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 798 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-200-900A-2

Query Match 34.8%; Score 536.5; DB 1; Length 798;
Best Local Similarity 39.2%; Pred. No. 6.5e-50;
Matches 111; Conservative 53; Mismatches 104; Indels 15; Gaps 6;

QY 1 MKLNTSAGNVDIYKLYHSDACSSKAVVSLRC-IACGVNLNSSROS-RIVGESALPG 57
DB 349 URLSTDSPTRTTYAOGSSGYSLRCKYVSSDCTTKIN---ATVGTGTSJGBEWP 404

DB 518 VNLNTAPNGSLI---LTPSQQLSDSLILQCNYSCKGLVTOEVSFKIVGSDSREGA 574

QY 58 WPMOVSILHVQNVAVCGGSITTPETVITAAHCYKELNPMHTAFAGILROSMFYGAGY 117

DB 575 WPMVVALYFDDQVCASIVSRDMLVSAHCYGRMBESKKAVALGLHMASTL---TSP 631

QY 118 QVE-----KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKVCPLNPGMLOPEOLCWI 172

DB 632 QLETLIDQIVNPHNKRKNDIAMMHLKAVTTDIOPLCLPEBNQVPPPGACSI 631

QY 173 SCMGATBEKKTSEVLNAKVLLEFORCNSRYVDNLITPMICAGFLQGNVDSQGS 232

DB 692 AGMGALIVQGSTADVLQEADVPLLSNEKCOQMPERN-ITENNVACAGYAGVDSQGS 750

QY 233 GGLVTSKNIMWLIGDTSMGSCAKAYRPGYGVNMTFTWY 275

DB 751 GGLPMQENNRWLLAGVTSFGYCALPNRPGYARVPRFTW 793

RESULT 13

PCT-US94-00616-2

Sequence 2, Application PC/TUS9400616

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE

NUMBER OF SEQUENCES: 33

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/00616

FILING DATE:

CLASSIFICATION:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 798 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-US94-00616-2

Query Match 34.8%; Score 536.5; DB 5; Length 798;
Best Local Similarity 39.2%; Pred. No. 6.5e-50;
Matches 111; Conservative 53; Mismatches 104; Indels 15; Gaps 6;

QY 1 MKLNTSAGNVDIYKLYHSDACSSKAVVSLRC-IACGVNLNSSROS-RIVGESALPG 57
DB 518 VNLNTAPNGSLI---LTPSQQLSDSLILQCNYSCKGLVTOEVSFKIVGSDSREGA 574

QY 58 WPMOVSILHVQNVAVCGGSITTPETVITAAHCYKELNPMHTAFAGILROSMFYGAGY 117
DB 575 WPMVVALYFDDQVCASIVSRDMLVSAHCYGRMBESKKAVALGLHMASTL---TSP 631

QY 118 QVE-----KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKVCPLNPGMLOPEOLCWI 172
DB 632 QLETLIDQIVNPHNKRKNDIAMMHLKAVTTDIOPLCLPEBNQVPPPGACSI 631

QY 173 SCMGATBEKKTSEVLNAKVLLEFORCNSRYVDNLITPMICAGFLQGNVDSQGS 232
DB 692 AGMGALIVQGSTADVLQEADVPLLSNEKCOQMPERN-ITENNVACAGYAGVDSQGS 750

QY 233 GGLVTSKNIMWLIGDTSMGSCAKAYRPGYGVNMTFTWY 275
DB 751 GGLPMQENNRWLLAGVTSFGYCALPNRPGYARVPRFTW 793

RESULT 14

US-09-820-002-2

Sequence 2, Application US/09820002

Patent No. 6482630


```

; GENERAL INFORMATION:
; APPLICANT: Gan, Weinlu
; APPLICANT: Ye, Jane
; APPLICANT: DiFrancesco, Valentina
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; FILE REFERENCE: CL001194
; CURRENT APPLICATION NUMBER: US/09/820,002
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 376
; TYPE: PRT
; ORGANISM: HUMAN
US-09-820-002-2

```

```

Query Match          33.9%; Score 522.5; DB 4; Length 376;
Best Local Similarity 37.4%; Pred. No. 7,5e-49;
Matches 108; Conservative 44; Mismatches 106; Indels 31; Gaps 5;

```

```

QY 22 CSSKA---VSLRCLACGVNLNSR-----QSRIVGESALPGAMPNQ 61
DB 77 CSSRANARVAGLSCSEMGLSDCPGRFLAALCQDCGRRRLPYDRITVGGRDTSIGRMPNQ 136
QY 62 VSLHQNVAHVCSSITTPRMTVTAACVCEKPLNPMHTAFAGILROSEFMFYAGATQVER 121
DB 137 VSLRYDGAHLCCGSLISGWLTAACPFERNRVLSEMRVFAGAVAOASP-HGIQLGVQA 195
QY 122 VISHPNY-----DSKTKNDIALMKLOKPLTFNDLVKPVCLPNPGMLOPEOLCMTISGW 175
DB 196 VVYHGGYLPFRPNSEENNDIALVHLSPLPTBYIQVCLPAAGALVDGKICTYTGW 255
QY 176 GATEBKGTSEVLNAKATLLIETORCNSRYVYDNLITPAMICAGFLQGVNDSQGGSGGX 235
DB 256 GNTQYGGQAGVLOEARVPILISNDVCNADFYGNQIKPKMFCAGYPEGGIDACQDSGGP 315
QY 236 LV---TSKNNTWMLIGDTSKSGCAKAYRPGYGVNMVFTDWIYRQMR 280
DB 316 FVCEDSISRTPRMRLGIVSGWGTGALAKPGVYTVSDFRFWIPOAIK 364

```

RESULT 15

```

US-09-820-002-4
; Sequence 4, Application US/09820002
; Patent No. 6482630
; GENERAL INFORMATION:
; APPLICANT: Gan, Weinlu
; APPLICANT: Ye, Jane
; APPLICANT: DiFrancesco, Valentina
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; FILE REFERENCE: CL001194
; CURRENT APPLICATION NUMBER: US/09/820,002
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 417
; TYPE: PRT
; ORGANISM: HUMAN
US-09-820-002-4

```

```

Query Match          33.7%; Score 520.5; DB 4; Length 417;
Best Local Similarity 38.7%; Pred. No. 1.4e-48;
Matches 104; Conservative 43; Mismatches 109; Indels 13; Gaps 4;
QY 22 CSSRAVSLRCLACGVNLNSRQSRIVGSEALPGAMPQVSLHQNVAHVCSSITTPEM 81

```

```

DB 140 CPRGRFLAALCQDCG---RRRLPYDRITVGGRDTSIGRMPQVSLRYDGAHLCCGSLISGDW 197
QY 82 IYVTAACVCEKPLNPMHTAFAGILROSEFMFYAGATQVERVISHPNY-----DSKTKN 135
DB 198 VLTAAHCFPERNRVLSRMRVFAGAVAOASP-HGIQLGVQAVVYHGGYLPDRDPNSENSN 256
QY 136 DIAMKLOKPLTFNDLVKPVCLPNPGMLOPEOLCMTISGKATSEKGTSEVLNAKATLL 195
DB 257 DIALVHLSPLPTBYIQVCLPAAGALVDGKICTYTGKNTQYGGQAGVLOEARVPI 316
QY 196 IETORCNSRYVYDNLITPAMICAGFLQGVNDSQGGSGGXLY---TSKNNTWMLIGDTS 251
DB 317 ISNDVCNADFYGNQIKPKMFCAGYPEGGIDACQDSGGPFCEDSISRTPRMRLGIVS 376
QY 252 WSGCAKAYRPGYGVNMVFTDWIYRQMR 280
DB 377 WGTGALAKPGVYTVSDFRFWIPOAIK 405

```

```

Search completed: August 1, 2003, 18:13:31
Job time : 21 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2003, 18:09:49 ; Search time 36 Seconds
(without alignments)

2028.580 Million cell updates/sec

Title: US-09-988-975A-1

Sequence: 1 MKNLTSAGNVDIYKKLYHSD.....YGVNVMVPTDWTYRQWRADG 283

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-----------|--------------------|
| 1 | 1540 | 99.8 | 492 | 4 Q96T73 | Q96T73 homo sapien |
| 2 | 1263.5 | 81.9 | 490 | 11 Q920K3 | Q920K3 rattus norv |
| 3 | 764.5 | 49.5 | 767 | 13 Q9DGR2 | Q9DGR2 xenopus lae |
| 4 | 695 | 45.0 | 453 | 11 Q8VDRO | Q8VDRO mus musculu |
| 5 | 694 | 45.0 | 453 | 11 Q8K1T0 | Q8K1T0 mus musculu |
| 6 | 595 | 38.6 | 471 | 11 Q8CFE0 | Q8CFE0 mus musculu |
| 7 | 588 | 38.1 | 537 | 4 Q9BYE1 | Q9BYE1 homo sapien |
| 8 | 587.5 | 38.1 | 371 | 11 Q8CJ16 | Q8CJ16 rattus norv |
| 9 | 587.5 | 37.9 | 445 | 11 Q8CJ17 | Q8CJ17 rattus norv |
| 10 | 585 | 37.9 | 581 | 4 Q9BYE2 | Q9BYE2 homo sapien |
| 11 | 572.5 | 37.1 | 455 | 11 Q8CDRO | Q8CDRO mus musculu |
| 12 | 568.5 | 36.2 | 405 | 4 Q9E6B6 | Q9E6B6 homo sapien |
| 13 | 553 | 35.1 | 435 | 11 Q8VCAS | Q8VCAS mus musculu |
| 14 | 547.5 | 35.8 | 643 | 6 Q97506 | Q97506 sus scrofa |
| 15 | 547.5 | 35.5 | 624 | 11 Q9DAT3 | Q9DAT3 mus musculu |
| 16 | 541.5 | 35.1 | 624 | 11 Q91Y47 | Q91Y47 mus musculu |

| | | | | | |
|----|-------|------|------|-----------|--------------------|
| 17 | 534 | 34.6 | 624 | 6 Q95ME7 | Q95ME7 oryctolagus |
| 18 | 530.5 | 34.4 | 638 | 11 Q8R0P5 | Q8R0P5 mus musculu |
| 19 | 522 | 33.8 | 777 | 11 Q8CAN9 | Q8CAN9 mus musculu |
| 20 | 520 | 33.7 | 310 | 11 Q9QY29 | Q9QY29 mus musculu |
| 21 | 514 | 33.3 | 310 | 11 Q91XC4 | Q91XC4 mus musculu |
| 22 | 512.5 | 33.2 | 422 | 4 Q8WVC1 | Q8WVC1 homo sapien |
| 23 | 511 | 33.1 | 799 | 11 Q9DB10 | Q9DB10 mus musculu |
| 24 | 509.5 | 33.0 | 855 | 11 Q9J1T7 | Q9J1T7 rattus norv |
| 25 | 507 | 32.9 | 439 | 11 Q8BHM9 | Q8BHM9 mus musculu |
| 26 | 506 | 32.8 | 417 | 11 Q8VHJ4 | Q8VHJ4 rattus norv |
| 27 | 504.5 | 32.7 | 812 | 11 Q91WJ5 | Q91WJ5 mus musculu |
| 28 | 503 | 32.6 | 417 | 11 Q8VDV1 | Q8VDV1 mus musculu |
| 29 | 503 | 32.6 | 417 | 11 Q8VHK8 | Q8VHK8 mus musculu |
| 30 | 502.5 | 32.6 | 812 | 11 Q9R0W3 | Q9R0W3 rattus norv |
| 31 | 500.5 | 32.4 | 329 | 6 Q9GJ10 | Q9GJ10 ovis aries |
| 32 | 500.5 | 32.4 | 331 | 11 Q8RI46 | Q8RI46 mus musculu |
| 33 | 500.5 | 32.4 | 389 | 13 Q9PVX7 | Q9PVX7 xenopus lae |
| 34 | 496 | 32.1 | 328 | 11 Q8BJR6 | Q8BJR6 mus musculu |
| 35 | 495.5 | 32.1 | 257 | 11 Q8BZ04 | Q8BZ04 mus musculu |
| 36 | 495.5 | 32.1 | 417 | 11 Q8BZ10 | Q8BZ10 mus musculu |
| 37 | 494.5 | 32.0 | 279 | 11 Q9QZ74 | Q9QZ74 rattus norv |
| 38 | 490 | 31.8 | 1524 | 13 Q91674 | Q91674 xenopus lae |
| 39 | 484 | 31.4 | 802 | 4 Q81UE2 | Q81UE2 homo sapien |
| 40 | 484 | 31.4 | 811 | 4 Q81U80 | Q81U80 homo sapien |
| 41 | 481.5 | 31.2 | 415 | 6 Q29015 | Q29015 sus sp. pre |
| 42 | 480.5 | 31.2 | 572 | 11 Q8BHK6 | Q8BHK6 mus musculu |
| 43 | 478 | 31.0 | 845 | 13 Q9DGR1 | Q9DGR1 xenopus lae |
| 44 | 476.5 | 30.9 | 416 | 11 Q8BZ30 | Q8BZ30 mus musculu |
| 45 | 476.5 | 30.9 | 416 | 11 Q8BZ13 | Q8BZ13 mus musculu |

ALIGNMENTS

RESULT 1
Q96T73 PRELIMINARY; PRT; 492 AA.
ID Q96T73
AC Q96T73;
DT 01-DEC-2001 (TRMBLrel. 19, Created)
DT 01-DEC-2001 (TRMBLrel. 19, last sequence update)
DT 01-MAR-2003 (TRMBLrel. 23, last annotation update)
DE Bcltheliasin.
GN TMPSR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
ON NCBI_TaxId=9606;
RX MEDLINE=21223025; PubMed=11322890;
RA Jachquinet E., Rao N.V., Rao G.V., Wang Z., Albertine K.H.,
RA Hoidal J.R.;
RT "Cloning and characterization of the cDNA and gene for human
RT epitheliasin";
RL Bnr. J. Biochem. 268:2687-2699(2001).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AF329454; AAK5359.1; -
DR HSSP; P00761; IAN1.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR001190; Src_receptor.
DR Pfam; PF00089; trypsin_1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00202; SR_1.
DR SMART; SM00202; TRYP_SPC_1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS0068; LDLRA_2; 1.
DR PROSITE; PS0287; SRCR_2; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HTS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.

KM Hydrolyase; Protease; Serine protease.
 SQ SEQUENCE 492 AA; 53863 MW; 3ABA755B276DADF CRC64;
 Query Match 99.8%; Score 1540; DB 4; Length 492;
 Best Local Similarity 99.6%; Pred. No. 2,6e-144;
 Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKLNTSAGNVDIYKLYHSDACSSKAVVSLRCIACGVNLNSRQRIYGGESALPGAMPW 60
 DB 210 MKLNTSAGNVDIYKLYHSDACSSKAVVSLRCIACGVNLNSRQRIYGGESALPGAMPW 269
 QY 61 QVSLHVQNVHYCGGSIITPEMIVTAHCVCKPLNPMHTAPAGILROSFMFYGAGYQVE 120
 DB 270 QVSLHVQNVHYCGGSIITPEMIVTAHCVCKPLNPMHTAPAGILROSFMFYGAGYQVE 329
 QY 121 KVISHPNYSKTKNNDIALMKLOKPLTENDLVKPYCLPNPGMMLQPEOLCWSMGATEE 180
 DB 330 KVISHPNYSKTKNNDIALMKLOKPLTENDLVKPYCLPNPGMMLQPEOLCWSMGATEE 389
 QY 181 KGKTSSEVLAANKVLLIFTRQNSRYVDNLITPAMICAGFLQGVNDSGCGSGGGLVTSK 240
 DB 390 KGKTSSEVLAANKVLLIFTRQNSRYVDNLITPAMICAGFLQGVNDSGCGSGGGLVTSK 449
 QY 241 NNIMWLLIGDTSWGGSCAKARYPGVYGNVFTDWTYRQBRADG 283
 DB 450 NNIMWLLIGDTSWGGSCAKARYPGVYGNVFTDWTYRQBRADG 492

RESULT 2
 0920K3 PRELIMINARY; PRT; 490 AA.
 AC 0920K3
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE TMRSS2.
 GN TMRSS2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tsubaki S.;
 RT "TMRSS2, Rat.";
 RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 DR EMBL; AB073550; BAB70683.1; -.
 DR HSSP; P00761; 1ANI.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR InterPro; IPR001190; Strc_receptor.
 DR Pfam; PF00057; ldl_recept_a; 1.
 DR Pfam; PF00089; ctrypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00192; LDLA; 1.
 DR SMART; SM00202; SR; 1.
 DR SMART; SM00020; TYP_SPC; 1.
 DR PROSITE; PS50068; LDLRA_2; 1.
 DR PROSITE; PS50287; SRCK_2; 1.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolyase; Protease; Serine protease.
 SQ SEQUENCE 490 AA; 53518 MW; 2BC691551CAC409A CRC64;
 Query Match 81.9%; Score 1263.5; DB 11; Length 490;
 Best Local Similarity 81.6%; Pred. No. 7.6e-117;
 Matches 230; Conservative 22; Mismatches 29; Indels 1; Gaps 1;

QY 1 MKLNTSAGNVDIYKLYHSDACSSKAVVSLRCIACGVNLNSRQRIYGGESALPGAMPW 60
 DB 209 MKLNTSAGNVDIYKLYHSDACSSKAVVSLRCIACGVNLNSRQRIYGGESALPGAMPW 267

QY 61 QVSLHVQNVHYCGGSIITPEMIVTAHCVCKPLNPMHTAPAGILROSFMFYGAGYQVE 120
 DB 268 QVSLHVQNVHYCGGSIITPEMIVTAHCVCKPLNPMHTAPAGILROSFMFYGAGYQVE 327
 QY 121 KVISHPNYSKTKNNDIALMKLOKPLTENDLVKPYCLPNPGMMLQPEOLCWSMGATEE 180
 DB 328 KVISHPNYSKTKNNDIALMKLOKPLTENDLVKPYCLPNPGMMLQPEOLCWSMGATEE 387
 QY 181 KGKTSSEVLAANKVLLIFTRQNSRYVDNLITPAMICAGFLQGVNDSGCGSGGGLVTSK 240
 DB 388 KGKTSSEVLAANKVLLIFTRQNSRYVDNLITPAMICAGFLQGVNDSGCGSGGGLVTSK 447
 QY 241 NNIMWLLIGDTSWGGSCAKARYPGVYGNVFTDWTYRQBRADG 282
 DB 448 NNIMWLLIGDTSWGGSCAKARYPGVYGNVFTDWTYRQBRADG 489

RESULT 3
 09DGR2 PRELIMINARY; PRT; 767 AA.
 AC 09DGR2
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Embryonic serine protease-2.
 GN XESP-2.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_Taxid=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=20363741; PubMed=10903452;
 RA Yamada K.; Takabatake T.; Takeshima K.;
 RT "Isolation and characterization of three novel serine protease genes from Xenopus laevis.";
 RL Gene 252:209-216(2000).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 DR EMBL; AB038497; BAB08217.1; -.
 DR HSSP; P00766; 1CHG.
 DR MEROPS; S01.049; -.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR Pfam; PF00057; ldl_recept_a; 4.
 DR Pfam; PF00089; ctrypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR SMART; SM00192; LDLA; 8.
 DR SMART; SM00020; TYP_SPC; 1.
 DR PROSITE; PS01209; LDLRA_1; 8.
 DR PROSITE; PS50068; LDLRA_2; 2.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolyase; Protease; Serine protease.
 SQ SEQUENCE 767 AA; 86001 MW; E0566A38796D56E CRC64;
 Query Match 49.5%; Score 764.5; DB 13; Length 767;
 Best Local Similarity 52.3%; Pred. No. 4.9e-67;
 Matches 146; Conservative 35; Mismatches 97; Indels 1; Gaps 1;

QY 2 KMLNTSAGNVDIYKLYHSDACSSKAVVSLRCIACGVNLNSRQRIYGGESALPGAMPW 61
 DB 487 KLYSGTMRSKFTTSYQSYSTCYSGNVSLHCLSCGVN-SNLSVSRIVGGTFANLGNMPW 545
 QY 62 VSLHVQNVHYCGGSIITPEMIVTAHCVCKPLNPMHTAPAGILROSFMFYGAGYQVE 121
 DB 546 VMLQGVITGLVCGGSIITPEMIVTAHCVGYSSASGKRVFVAGTILKPSYNAASAVFVER 605
 QY 122 VISHPNYSKTKNNDIALMKLOKPLTENDLVKPYCLPNPGMMLQPEOLCWSMGATEE 181

```
Db      606 IIVHGYSYYNDNIALMKRDEITFGYTTQPVCLPNSGMFWAGTITTMISGSGTYSYG 665
Qy      182 GKTSEVLAAKVLLETQRCNSRYYDNLITPAMICAGFLQGVNDSGQDSGGXLTSTKN 241
Db      666 GSVETIYLYAAIPLIDSNVQNSYVNGCITSSMICAGTISGVDTCQDGGGGLVKNKN 725
Qy      242 NIWMLIGDTSMGSGCAKAYRPGVYGVNVFTDVIYRQR 280
Db      726 GTWMLVGDTSWGDGCABANKPGVYGVNVFTFLMIYSQMR 764

RESULT 4
Q8VDE0
ID 08VDE0 PRELIMINARY; PRT; 453 AA.
AC 08VDE0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE TMPS33 protein.
GN TMPS33.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Guipont M., Scamuffa N., Scott H.S., Rossier C., Antonarakis S.E.,
RT "Isolation and characterization of the mouse Tmp33 gene."
RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AJ300738; CAC83350.1; -
DR HSSP; P00761; IAN1.
DR MED; MG1:2155445; Tmp33.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR001190; Src_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLA; 1.
DR SMART; SM00202; Tryp_Spc; 1.
DR PROSITE; PS01209; LDLA_1; 1.
DR PROSITE; PS50068; LDLA_2; 1.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 453 AA; 49529 MW; 2185697DC8781BD3 CRC64;
```

```
Db      419 SFGIGCAEVNKKPGVYTRITSPFLDMIHQJERD 450

RESULT 5
Q8X1T0
ID 08X1T0 PRELIMINARY; PRT; 453 AA.
AC 08X1T0;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Type II transmembrane serine protease.
GN TMPS33.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Guipont M., Scamuffa N., Scott H.S., Antonarakis S.E.,
RT "Isolation of the mouse Tmp33 genomic DNA sequence."
RL Submitted (JAN-2002) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AJ29216; CAD2137.1; -
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR001190; Src_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLA; 1.
DR SMART; SM00202; SR; 1.
DR PROSITE; PS01209; LDLA_1; 1.
DR PROSITE; PS50068; LDLA_2; 1.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease; Transmembrane.
SQ SEQUENCE 453 AA; 49491 MW; 1ABC8F10AE631EF6 CRC64;
```

Query Match 45.0%; Score 694; DB 11; Length 453;
Best Local Similarity 49.6%; Pred. No. 2.5e-60;
Matches 135; Conservative 39; Mismatches 92; Indels 6; Gaps 4;

```
Qy      12 IYKLYHSDAGSSKAVSIRCLACGVNANSSRQSRIVGGSALPGAMPQVSLAVQNVHY 71
Db      184 LHSVYMEGCTSGHVVTLKCSACGTRTGY--PRIVGMSLSLTQMPQVSLQFOGYHL 241
Qy      72 CGGSITPEWITVAHCEKPLNPMHTAFAGILRQSFYGA-GYQEKYISHNNYS 130
Db      242 CGGSITITPLMIVTAHCV-YDLHYKSWTVQGLV--SLMSPVPSHLVEKTIYHSKYP 298
Qy      131 KTKNDIALMLQKQLFENDLVKPCLPNPGMLOPEOLCWSIGGATEEKGKTSSEVLNA 190
Db      299 KRLGNDIALMLQKSELTGDETIOPICLPNSBNPFDGKLCWISGNGATEDGDADSPVLNH 358
Qy      191 AKVLLIEFQRCNSRYYDNLITPAMICAGFLQGVNDSGQDSGGXLTSTKNIMWLGDT 250
Db      359 AAVPLISKICNHRDVGIIISPSMLCAGYLKGVDSGQDSGGGLVLCQERRLMKLVGAT 418
Qy      251 SWGSCAKAYRPGVYGVNVFTDVIYRQRAD 282
Db      419 SFGIGCAEVNKKPGVYTRITSPFLDMIHQJERD 450

RESULT 6
Q8CFE0
ID 08CFE0 PRELIMINARY; PRT; 471 AA.
AC 08CFE0;
```

```
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DB Similar to mosaic serine protease (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Breast tumor;
RA Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC042878; AAH42878.1; -.
KM Protease.
FT NON_TER
SQ SEQUENCE 471 AA; 52535 MW; ED58CFE6B7C3BC4 CRC64;

Query Match 38.6%; Score 595; DB 11; Length 471;
Best Local Similarity 43.0%; Pred. No. 1.8e-50;
Matches 120; Conservative 43; Mismatches 102; Indels 14; Gaps 6;

QY 9 NVDIYKLYHSDACSKAVSLRCLACGVNLNLSRQSHIVGSGSLPGAMPQVSLHYQN 68
DB NTIIQESLYRSQ-CPSRRYVSLQCSHGCLR--ANTGRIVGALNSESMPQVSLHFGT 251
QY 69 VHVCGGSIITPEWVITAAHCV---EKPLNPMWMTAFAGILRQSFMYGAGYQVEKYIS 124
DB THICGGLTLDQWVLTAAHCFVTRKLELG---WKVYAGTSNLHQLPEAA--SISQILI 306
QY 125 HPNVDKTKRNDIALMKLQKPLTFNDLVKPVCLPMPGMLOPEQICWISGMKATEBKX 183
DB NGNTDEDDDDYDIALIRLSKPLTSAHHPACLPMHGOTFSLNFCWTGFKETETDEK 366
QY 184 TSEVLNAKVLIIETORCNSRYVDNLITPMICAGPLQGVNDSQSGSGGLVYSKNNI 243
DB TSPFLREVQVNLIDPKKNDLVYDSYLTFRMCMAGDLHGGRDSQSGSGGLVYCEQNNR 426
QY 244 WMLIGDTWSGSCAKAVRPGYGVNMFVTWYROMAD 282
DB WYLAGVTSWGTGCGQKRNKPGYTTKTVLPMYISQSEB 465

RESULT 7
Q9BYEI PRELIMINARY; PRT; 537 AA.
ID Q9BYEI
AC Q9BYEI;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Mosaic serine protease.
GN MSPS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RC MEDLINE=21167393; PubMed=11267681;
RA Kim D.R., Sharmir S., Inoue M., Kido H.;
RT "Cloning and expression of novel mosaic serine proteases with and
without a transmembrane domain from human lung.";
RL Biochim. Biophys. Acta 1518:204-209 (2001).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL: AB048797; BAB39742.1; -.
DR HSSP: P00763; IDPO.
DR MEROPS: S01.067; -.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR002172; LDL_receptor_A.
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR001190; Srcr_receptor.
DR Pfam: PF00057; Idl_recept_a; 1.
```

```
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00192; LDLr; 1.
DR SMART: SM00202; SR; 1.
DR SMART: SM00200; Tryp_Spc; 1.
DR PROSITE: PS50287; SRCR_2; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SSR; 1.
DR Hydrolase; Protease; Serine protease.
SQ SEQUENCE 537 AA; 58102 MW; A39FAE8616DAECF CRC64;

Query Match 38.1%; Score 588; DB 4; Length 537;
Best Local Similarity 42.7%; Pred. No. 1.1e-49;
Matches 119; Conservative 41; Mismatches 105; Indels 14; Gaps 6;

QY 9 NVDIYKLYHSDACSKAVSLRCLACGVNLNLSRQSHIVGSGSLPGAMPQVSLHYQN 68
DB NTIIQESL-HSHKCSQRYISLQCSHGCLR--ANTGRIVGALASDSKMPQVSLHFGT 317
QY 69 VHVCGGSIITPEWVITAAHCV---EKPLNPMWMTAFAGILRQSFMYGAGYQVEKYIS 124
DB THICGGLTLDQWVLTAAHCFVTRKLELG---WKVYAGTSNLHQLPEAA--SISQILI 372
QY 125 HPNVDKTKRNDIALMKLQKPLTFNDLVKPVCLPMPGMLOPEQICWISGMKATEBKX 183
DB NSNTDEDDDDYDIALIRLSKPLTSAHHPACLPMHGOTFSLNFCWTGFKETETDDK 432
QY 184 TSEVLNAKVLIIETORCNSRYVDNLITPMICAGPLQGVNDSQSGSGGLVYSKNNI 243
DB TSPFLREVQVNLIDPKKNDLVYDSYLTFRMCMAGDLHGGRDSQSGSGGLVYCEQNNR 492
QY 244 WMLIGDTWSGSCAKAVRPGYGVNMFVTWYROMAD 282
DB WYLAGVTSWGTGCGQKRNKPGYTTKTVLPMYISQSEB 531

RESULT 8
Q8CJ16 PRELIMINARY; PRT; 371 AA.
ID Q8CJ16
AC Q8CJ16;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Adrenal mitochondrial protease short variant.
GN AMP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEBH;
RA Omer S., Bicknell A.B., Lowry P.J.;
RT "Identification of a rat adrenal mitochondrial protease.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF537039; AAN06758.1; -.
KM Protease.
SQ SEQUENCE 371 AA; 40694 MW; 89A64081D9A1FE26 CRC64;

Query Match 38.1%; Score 587.5; DB 11; Length 371;
Best Local Similarity 42.7%; Pred. No. 7.4e-50;
Matches 111; Conservative 39; Mismatches 105; Indels 5; Gaps 3;

QY 19 SDACSKAVSLRCLACGVNLNLSRQSHIVGSGSLPGAMPQVSLHYQNHYVCGGSIIT 78
DB STNCPGKIVSLKCSBCAR--PLASRIVGQAVASGRMPQVSLHFGTGGSVLA 165
QY 79 PEWITTAHCVK-PLNPMWMTAFAGILRQSFMYGAGYQVEKYISHPNYDSKTKNDI 137
DB PYWVITAAHCYVSLRSLHSRVAHGVSHSAVRHQGTWVEKILPHPLVSAQNHDDV 225
QY 138 ALMKLQKPLTFNDLVKPVCLPMPGMLOPEQICWISGMKATEBKXGKTSSEVLNAKVLII 196
```

```

Db 226 ALIQRTPIINFSDDTSAVCLPAKEQHFPQSGQCVSGWGHGTHSDTHSSDTLQDTWVPLL 285
Qy 197 ETORCNSRYVYDNLITPAMICAGFLQGNVDSGCGSGKLYTSKNNIMWLIGDTSGSGC 256
Db 286 STDLCNSSSCMTSGALTHRLCAGYLDGRADACGCGSGGLVCPSPGDTWHLVGVVSGRG 345
Qy 257 AKAYRPGVYGVNVMTDWTY 276
Db 346 AEPNRPGYAKVAEFLDWH 439

```

RESULT 9

```

08CJ17 PRELIMINARY; PRT; 445 AA.
ID 08CJ17
AC 08CJ17
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DE Adrenal mitochondrial protease long variant.
GN AMP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEBH;
RA Omer S., Bicknell A.B., Lowry P.J.;
RT "Identification of a rat adrenal mitochondrial protease."
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF537098; AAN06757.1; -.
KM
SQ SEQUENCE 445 AA; 48440 MW; B33F56D8372ED988 CRC64;

```

```

Query Match 38.1%; Score 587.5; DB 11; Length 445;
Best Local Similarity 42.7%; Pred. No. 9.3e-50;
Matches 111; Conservative 39; Mismatches 105; Indels 5; Gaps 3;

```

```

Qy 19 SDACSKAVVSLRCLACGVNLNSRQSRVIGESALPGAMPQVSLHVNHVCGSIIT 78
Db 183 STNCPSGRIVSLKCECGAR---PLASRIVGQAVASGRMPQASVMLSGSRHTCGGSVLA 239
Qy 79 PEWITAHCEYK-PLANNPMTAFAGILROSPFPGAGYQVEKVISHNYSKTDNDI 137
Db 240 PYWVTAHCHVSPFLSRSLRSHVAGVSHSAVHOQGTWEKIIHPFLYSANQNDHYV 299
Qy 138 ALMKLQKPLTNDLVKPYCLPNPQMMLOPEQLCMTSGMGATE-EKGTSEVLAIAKVLII 196
Db 300 ALLQRTPIINFSDDTSAVCLPAKEQHFPQSGQCVSGWGHGTHSDTHSSDTLQDTWVPLL 359
Qy 197 ETORCNSRYVYDNLITPAMICAGFLQGNVDSGCGSGKLYTSKNNIMWLIGDTSGSGC 256
Db 360 STDLCNSSSCMTSGALTHRLCAGYLDGRADACGCGSGGLVCPSPGDTWHLVGVVSGRG 419
Qy 257 AKAYRPGVYGVNVMTDWTY 276
Db 420 AEPNRPGYAKVAEFLDWH 439

```

RESULT 10

```

09BYE2 PRELIMINARY; PRT; 581 AA.
ID 09BYE2
AC 09BYE2
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Membrane-type mosaic serine protease.
GN MSL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=21167393; PubMed=11267681;
RA Kim D.R., Sharmila S., Inoue M., Kido H.;
RT "Cloning and expression of novel mosaic serine proteases with and
RT without a transmembrane domain from human lung."
RL Blochm. Biophys. Acta 1518:204-209 (2001).
DR EMBL; AB048796; BAB39741.1; -.
DR HSP; P00763; IDPO.
DR MEROPS; S01.087; -.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR001190; Src_receptor.
DR Pfam; PF00057; Id1_recept_a; 1.
DR SMART; SM00089; trypsin; 1.
DR SMART; SM00192; Ldla; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00287; SRCR_2; 1.
DR PROSITE; PS0240; TRYPsin_DOM; 1.
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SRR; 1.
KM Hydrolyase; Protease; Serine protease.
SQ SEQUENCE 581 AA; 62689 MW; 4DABR24D7D5BA44 CRC64;

```

```

Query Match 37.9%; Score 585; DB 4; Length 581;
Best Local Similarity 43.1%; Pred. No. 2.3e-49;
Matches 119; Conservative 39; Mismatches 104; Indels 14; Gaps 6;

```

```

Qy 9 NVDIYKGLYSHDACGSKAVVSLRCLACGVNLNSRQSRVIGESALPGAMPQVSLHVN 68
Db 287 NSTTSESL-HRSHCPQRYSISLQCSHGRL---AMGRIVGALADSDSKPQVSLHFGT 342
Qy 69 VHVCGSIITPEWITAHCEYK---EKPLNNPMTAFAGILROSPFPGAGYQVEKVIS 124
Db 343 THICGTLIDAGWVTLPAHCFYTRKVLFG---WVYAGTSNLHQPEAA--SIARII 397
Qy 125 HENVDSKTDNDIALMKLQKPLTNDLVKPYCLPNPQMMLOPEQLCMTSGMGATEE-KGR 183
Db 398 NSNTYDEEDVDYDIALMRSLKPLTLSAHHPACLPWNGQTFSLNFTCMITGFTRETDK 457
Qy 184 TSEVLAIAKVLIIETORCNSRYVYDNLITPAMICAGFLQGNVDSGCGSGKLYTSKNNI 243
Db 458 TSPFLAEVYVNLIDPKKCDYLYDSYLTFRMWCAGDLHGGRDSGCGSGGLVCGQNNR 517
Qy 244 WMLIGDTSGSGCAKAYRPGVYGVNVMTDWTYRQW 279
Db 518 WTLAAGVTSMTGCGGRNRPQVITKTEVLPWITSK 553

```

RESULT 11

```

08CDRO PRELIMINARY; PRT; 455 AA.
ID 08CDRO
AC 08CDRO
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Transmembrane protease.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=14466851;
RA The FANTOM Consortium;
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT Nature 420:563-573 (2002).";
DR EMBL; AK029714; BAC28577.1; -.

```

SEQUENCE 455 AA; 49669 MW; BE22EB2B7503C74B CRC64;

Query Match 37.1%; Score 572.5; DB 11; Length 455;
Best Local Similarity 42.0%; Pred. No. 3e-48;
Matches 108; Conservative 39; Mismatches 105; Indels 5; Gaps 3;

QY 22 CSSKAVVSLRACIACGVNLNSRQSRIVGSESLPGAMPVQVSLHVNQVNVCGSIT 81
DB 196 CPSRRIVSLKCEGAR--PLASRIVGQAVASGRWQMSVWLGSRHTGASVLA 252
QY 82 IVTAHCEK-PLNNPMTAFAGILRQSFMEYAGYQVEKYISHPNYSKTKNNDI 140
DB 253 VVTAHCHYSPRLSRISMRVHAGVSHGAVRQHGTMVEKTIPLPLVSAQNDVD 312
QY 141 KLQCEPLTENDLVKVCCLPFGMLQPEQLCWSGATE-EKQKTSVLAQKATL 199
DB 313 QLRTPINFSDTVGAVCLPAKEQHFPWQSCWVSQGHGTHDPSTHTSDTLQDT 372
QY 200 RCNRRVYDNLITPAMICAGFLQGVNDSQGDGGLVTSKNNIMWLIGDTS 259
DB 373 LCNSCWSGALTRHMLCAGLIDGADACQGDGGLVCPGSDTWHLVGVWSGR 432
QY 260 YRPGVGVNVFTDMY 276
DB 433 NRPGVAKVAEFLDMH 449

RESULT 12

Q96E86 PRELIMINARY; PRT; 405 AA.

AC 096E86;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Similar to transmembrane protease, serine 4 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Ruteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN 12
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Strausberg R.;
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; BC012752; AAH12752.1; -.
DR HSSP; P00761; IAN1.
DR MEROPS; S01.034; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001172; LDL_receptor_A.
DR InterPro; IPR001254; Ser_protase_Try.
DR InterPro; IPR001190; Strc_receptor.
DR Pfam; PF00057; Idl_recept_a; 1.
DR Pfam; PF00089; trypsin_1; -.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDla; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00202; Tryp_Spc; 1.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease; Transmembrane.
FT NON TER 1
SQ SEQUENCE 405 AA; 44474 MW; 951ACD52D9D48E04 CRC64;

Query Match 36.2%; Score 558.5; DB 4; Length 405;
Best Local Similarity 43.9%; Pred. No. 6.3e-47;
Matches 118; Conservative 41; Mismatches 93; Indels 17; Gaps 8;

QY 19 SDACSKAVVSLRACIACGVNLNSRQSRIVGSESLPGAMPVQVSLHVNQVNVCGSIT 78
DB 148 SGRCLSGVSLVRLCLDCKSL---KTRPVGVGEVLPVDSWPMQVSIQVKKQHVCGSIT 204

QY 79 PENTVAHCEKPLNNPMTAFAGILR-QSFMEYAGYQVEKY---SHPNYSKTKN 134
DB 205 PHTVTAHCHCRKA-TDFNMKVRKRGSKLSP----PSLAAKIILIFNPMY--PKD 256
QY 135 NDIALMKLQCEPLTENDLVKVCCLPFGMLQPEQLCWSGATEEK-GKTSVLAQK 193
DB 257 NDIALMKLQCEPLTENDLVKVCCLPFGMLQPEQLCWSGATEEK-GKTSVLAQK 316
QY 194 LIETQRNRRVYDNLITPAMICAGFLQGVNDSQGDGGLVTSKNNIMWLIGDTS 253
DB 317 QVIDSTRCANADAVGSEVTERKMCAGIPEGVDTCQGDGGLVQSDQ-MHVVGIVSWG 375
QY 254 SGCAKAVRPGVGVNVFTDMY 282
DB 376 YGCGPSTPGVYTKVSAVLAIVYWKAR 404

RESULT 13

Q8VCAS PRELIMINARY; PRT; 435 AA.

AC 08VCAS;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Similar to transmembrane protease, serine 4 (Channel-activating
protease 2).
GN TMPSR54.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Ruteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 13
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (Jan-2002) to the EMBL/GenBank/DBJ databases.
RN 12
RP SEQUENCE FROM N.A.
RX MEDLINE; 22144321; PubMed; 12149280;
RA Vuagniaux G., Vallet V., Jaeger N.F., Hummler B., Rossier B.C.;
RT "Synergistic Activation of ENaC by Three Membrane-bound Channel-
activating Serine Proteases (mCAP1, mCAP2, and mCAP3) and Serum- and
Glucocorticoid-regulated Kinase (SGK1) in Xenopus Oocytes.";
RL J. Gen. Physiol. 120:191-201(2002).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; BC021368; AAH21368.1; -.
DR EMBL; AY043240; AAK65307.1; -.
DR HSSP; P00761; IAN1.
DR MEROPS; S01.034; -.
DR MGD; MGI:2384877; Tmpres4.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001172; LDL_receptor_A.
DR InterPro; IPR001254; Ser_protase_Try.
DR InterPro; IPR001190; Strc_receptor.
DR Pfam; PF00057; Idl_recept_a; 1.
DR Pfam; PF00089; trypsin_1; -.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease; Transmembrane.
FT NON TER 1
SQ SEQUENCE 435 AA; 47495 MW; DC52BA5A43E01369 CRC64;

Query Match 36.1%; Score 557.5; DB 11; Length 435;
Best Local Similarity 43.8%; Pred. No. 8.7e-47;
Matches 117; Conservative 41; Mismatches 96; Indels 13; Gaps 7;

QY 19 SDACSKAVVSLRACIACGVNLNSRQSRIVGSESLPGAMPVQVSLHVNQVNVCGSIT 78
DB 178 SRSCLSGVSLVRLCLDCKSL---KTRPVGVGEVLPVDSWPMQVSIQVKKQHVCGSIT 234

```

QY 79 PEMITAAHCEKPLN-NPMTATAPAGILROSEMTYAGYOYEKV-ISHPNIDSKTYKND 136
DB 235 PHMITAAHCEKRYLDVSSWKRVRAGSNIIIGNS-----PSLPVAKIPIAEPN-FLYPERKD 288
QY 137 IALKMLQKPLTFNDLVKPCVCLPNPGMLOPBOLCWISGWAITEEK-GKTSVTLNNAKUTL 195
DB 289 IALVLOMPLTFSSGVRPLCLPSPDEVLPAFPVWVIGWFTFENGKRSMDLLOKASVY 348
QY 196 IETORCNSRYVDNLTTPMICAGFLQGNVDSQSGSGGLVTSKNNIMWLIGDTSMGSG 255
DB 349 IDSTRCNMADVAGEVTAEMLCAGTPOGSKDTCCGDSGGLPMTHSK-NQVVGISWKGK 407
QY 256 CAKAYRPGVYGNVVFPTDWTYROMRAD 282
DB 408 CGGPSTPGVYTKVTAYLWVYVNRKSE 434

RESULT 14
ID 097506 PRELIMINARY; PRT; 643 AA.
AC 097506;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Kallikrein.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
OX NCBI_TaxID=9623;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=plasma;
RX MEDLINE=20413446; PubMed=10954859;
RA Kimura A., Kihara T., Okimura H., Hamabata T., Ohnishi J.,
RA Moriyama A., Takahashi K., Takahashi T.,
RT "Identification of porcine follipain as plasma kallikrein, and its
RT possible involvement in the production of bradykinin within the
RT follicles of porcine ovaries."
RL Mol. Reprod. Dev. 57:79-87(2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AB022425; BAA37147.1; -.
DR HSSP; P00766; 1CHG.
DR MEROPS; S01.212; -.
DR InterPro; IPR000177; Apple.
DR InterPro; IPR000314; Chymotrypsin.
DR InterPro; IPR003014; PAN.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00024; PAN; 4.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00005; APPLDOMAIN.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00223; APPLE; 4.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00495; APPLE; 4.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 643 AA; 72227 MW; AFR2923E33CB80A CRC64;

Query March 35.8%; Score 553; DB 6; Length 643;
Best Local Similarity 41.5%; Pred. No. 4e-46; Indels 26; Gaps 7;
Matches 112; Conservative 43; Mismatches 89;
QY 18 HSDACSKAVVSLICAGVNLNRSRIVGSGSALPGAMPQVS-----LTFQNVHYCG 73
DB 388 HS-ACATKA-----NTRIIGTDSPLGHWVQVSLQAKLRAGN-HLGG 428
QY 74 GSITPEWITAAACEKPLNPMHTATAPAGILROSEMTYAGYQY-QVEKVISHPYDSKT 132
DB 429 GSIIHQWVLTAAHCFD-GLSLPDIWYIGGLINISSETKPEFSQVKEIITHQYKILE 487
QY 133 KNDIALMLKQKPLTFNDLVKPCVCLPNPGMLOPBOLCWISGWAITEEKGKTSVTLNNAK 192

```

```

DB 488 SGHDIALMLKLETPNTDTPQKPCVLRDDTWNVTYNTCWATGWFTEKKEIGNILQKN 547
QY 193 VLLIETORCNSRYVDNLTTPMICAGFLQGNVDSQSGSGGLVTSKNNIMWLIGDTSM 252
DB 548 IPLVNSBQCKSY-RDHLKSKQMICAGYREGGADKGSBGLVCKYKNGIMWLIGDTSM 606
QY 253 GSGCAYRPGVYGNVVFPTDWTYROMRAD 282
DB 607 GSGCAYRPGVYTKVTAYLWVYVNRKSE 434

RESULT 15
ID 09DAT3 PRELIMINARY; PRT; 624 AA.
AC 09DAT3;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE 1600027G01RIK protein (RIKEN cDNA 1600027G01 gene).
GN 1600027G01RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=placenta;
RX MEDLINE=21085660; PubMed=11217651;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Araiwa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamahata I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Pleischmann W., Gaasterland T., Glass C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Mashio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boileau D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guelincich S., Hill D., Hofmann M., Hume D.A., Kamaya M., Lee N.H.,
RA Lyone P., Marchionni L., Mashima Y., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Saeki H., Sato K., Schenbach C., Seta T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohseki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=liver;
RX Strusberg R.;
RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE EMBL/Genbank/DBJ databases.
DR EMBL; AK005546; BAB24114.1; -.
DR EMBL; BC019485; AAH19485.1; -.
DR HSSP; P00750; 1RTF.
DR MEROPS; S01.213; -.
DR MGD; MGI:1919281; 1600027G01RIK.
DR InterPro; IPR000177; Apple.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR003014; PAN.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00024; PAN; 4.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00005; APPLDOMAIN.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00223; APPLE; 4.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00495; APPLE; 3.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.

```


DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 624 AA; 69788 MW; 0EEDEB5C6009E97 CRC64;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: August 1, 2003, 18:12:44 / Search time 52 Seconds

(Without alignments)
646.327 Million cell updates/sec

Title: US-09-988-975A-1

Perfect score: 1543
Sequence: 1 MGLNTSAGNVNDIYKKLYHSD.....VYGNVWVFTDWIRQMRADG 283

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Published Applications AA.*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-------------------|-------------------|
| 1 | 1541 | 99.9 | 283 | US-09-988-975A-1 | Sequence 1, Appli |
| 2 | 1541 | 99.9 | 283 | US-10-235-699-1 | Sequence 1, Appli |
| 3 | 1540 | 99.8 | 393 | US-09-759-143-934 | Sequence 934, App |
| 4 | 1540 | 99.8 | 393 | US-09-780-669-934 | Sequence 934, App |
| 5 | 1540 | 99.8 | 393 | US-09-822-827-934 | Sequence 934, App |
| 6 | 1540 | 99.8 | 393 | US-09-822-827-934 | Sequence 934, App |
| 7 | 1540 | 99.8 | 393 | US-09-895-793-934 | Sequence 934, App |
| 8 | 1540 | 99.8 | 393 | US-09-895-814-934 | Sequence 934, App |
| 9 | 1540 | 99.8 | 492 | US-10-012-896-932 | Sequence 932, App |
| 10 | 1540 | 99.8 | 492 | US-09-759-143-932 | Sequence 932, App |
| 11 | 1540 | 99.8 | 492 | US-09-780-669-932 | Sequence 932, App |
| 12 | 1540 | 99.8 | 492 | US-09-822-827-932 | Sequence 932, App |
| 13 | 1540 | 99.8 | 492 | US-09-822-827-932 | Sequence 932, App |
| 14 | 1540 | 99.8 | 492 | US-09-895-814-932 | Sequence 932, App |
| 15 | 1540 | 99.8 | 492 | US-10-012-896-932 | Sequence 932, App |
| | | | | US-10-205-823-415 | Sequence 415, App |

| | | | | | | |
|----|-------|------|-----|----|-------------------|--------------------|
| 16 | 1535 | 99.5 | 384 | 10 | US-09-981-353-23 | Sequence 23, Appl |
| 17 | 1522 | 98.6 | 492 | 9 | US-09-759-143-895 | Sequence 895, App |
| 18 | 1522 | 98.6 | 492 | 9 | US-09-780-669-895 | Sequence 895, App |
| 19 | 1522 | 98.6 | 492 | 9 | US-09-822-827-895 | Sequence 895, App |
| 20 | 1522 | 98.6 | 492 | 9 | US-09-822-827-895 | Sequence 895, App |
| 21 | 1522 | 98.6 | 492 | 10 | US-09-895-793-895 | Sequence 895, App |
| 22 | 1522 | 98.6 | 492 | 10 | US-09-895-814-895 | Sequence 895, App |
| 23 | 1522 | 98.6 | 492 | 11 | US-09-776-191-70 | Sequence 70, Appl |
| 24 | 1522 | 98.6 | 492 | 14 | US-10-012-896-895 | Sequence 895, App |
| 25 | 1254 | 81.3 | 229 | 11 | US-09-898-837A-53 | Sequence 53, Appl |
| 26 | 778.5 | 50.5 | 418 | 15 | US-10-177-661-6 | Sequence 6, Appl1 |
| 27 | 696 | 45.1 | 453 | 10 | US-09-978-192A-69 | Sequence 69, Appl1 |
| 28 | 696 | 45.1 | 453 | 10 | US-09-978-192A-69 | Sequence 69, Appl1 |
| 29 | 696 | 45.1 | 453 | 10 | US-09-978-192A-69 | Sequence 69, Appl1 |
| 30 | 696 | 45.1 | 453 | 10 | US-09-978-192A-69 | Sequence 69, Appl1 |
| 31 | 696 | 45.1 | 453 | 11 | US-09-978-189-69 | Sequence 69, Appl1 |
| 32 | 696 | 45.1 | 453 | 11 | US-09-978-189-69 | Sequence 69, Appl1 |
| 33 | 696 | 45.1 | 453 | 11 | US-09-978-189-69 | Sequence 69, Appl1 |
| 34 | 696 | 45.1 | 453 | 11 | US-09-978-189-69 | Sequence 69, Appl1 |
| 35 | 696 | 45.1 | 453 | 11 | US-09-978-189-69 | Sequence 69, Appl1 |
| 36 | 696 | 45.1 | 453 | 11 | US-09-978-189-69 | Sequence 69, Appl1 |
| 37 | 696 | 45.1 | 453 | 11 | US-09-978-189-69 | Sequence 69, Appl1 |
| 38 | 696 | 45.1 | 453 | 11 | US-09-978-189-69 | Sequence 69, Appl1 |
| 39 | 696 | 45.1 | 453 | 11 | US-09-978-189-69 | Sequence 69, Appl1 |
| 40 | 696 | 45.1 | 453 | 11 | US-09-978-189-69 | Sequence 69, Appl1 |
| 41 | 696 | 45.1 | 453 | 11 | US-09-978-189-69 | Sequence 69, Appl1 |
| 42 | 696 | 45.1 | 453 | 11 | US-09-978-189-69 | Sequence 69, Appl1 |
| 43 | 696 | 45.1 | 453 | 11 | US-09-978-189-69 | Sequence 69, Appl1 |
| 44 | 696 | 45.1 | 453 | 11 | US-09-978-189-69 | Sequence 69, Appl1 |
| 45 | 696 | 45.1 | 453 | 11 | US-09-978-189-69 | Sequence 69, Appl1 |

ALIGNMENTS

RESULT 1
US-09-988-975A-1
Sequence 1, Application US/09988975A
Patent No. US20020119531A1
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
TITLE OF INVENTION: PROSTATE-ASSOCIATED PROTEASE ANTIBODY
FILE REFERENCE: PP-0227-2 CIP
CURRENT APPLICATION NUMBER: US/09/988,975A
CURRENT FILING DATE: 2001-11-19
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PERL Program
SEQ ID NO 1
LENGTH: 283
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20020119531A1 556016
FEATURE:
NAME/KEY: unsure
LOCATION: 235
OTHER INFORMATION: unknown or other
US-09-988-975A-1

Query Match 99.9%; Score 1541; DB 10; Length 283;

Best Local Similarity 100.0%; Pred. No. 2.5e-147; Indels 0; Gaps 0;

Matches 283; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLNTSAGNVNDIYKKLYHSDACSKAVVSLRICACGVNLSNRSHVIGESALPGAMPW 60
DB 1 MGLNTSAGNVNDIYKKLYHSDACSKAVVSLRICACGVNLSNRSHVIGESALPGAMPW 60
QY 61 QVSLHVNHHVGGSSITIPWTVAACHCEKRLANPMHTATPAGIIRSGFMFGAGYQVE 120
DB 61 QVSLHVNHHVGGSSITIPWTVAACHCEKRLANPMHTATPAGIIRSGFMFGAGYQVE 120

Qy 121 KVISHPNDYDKTKNDIALMKLOKPLTFNDLVKPVCLPMPGMMLOPEOLCMTISGMGATBE 180
Db 121 KVISHPNDYDKTKNDIALMKLOKPLTFNDLVKPVCLPMPGMMLOPEOLCMTISGMGATBE 180
Qy 181 KGTSEVINAARVLLIETORCNSRYVDNLITPAMI CAGFLQGNVDS CGDSGGLVTSK 240
Db 181 KGTSEVINAARVLLIETORCNSRYVDNLITPAMI CAGFLQGNVDS CGDSGGLVTSK 240
Qy 241 NNIMWLIGDTSWGSACAKAYRPGYGNVWVFTDWTYRORADG 283
Db 241 NNIMWLIGDTSWGSACAKAYRPGYGNVWVFTDWTYRORADG 283

RESULT 2

US-10-235-699-1
; Sequence 1, Application US/10235699
; Publication No. US20030103981A1
; GENERAL INFORMATION:
; APPLICANT: Spacelake, Kimberly M.
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti G.
; TITLE OF INVENTION: METHOD OF USE OF A PROSTATE-ASSOCIATED PROTEASE IN THE DIAGNOSIS
; FILE REFERENCE: PV-0010 CIP
; CURRENT APPLICATION NUMBER: US/10/235,699
; PRIOR FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: 09/988,975
; PRIOR FILING DATE: 11/19/01
; PRIOR APPLICATION NUMBER: 09/478,957
; PRIOR FILING DATE: 01/07/00
; PRIOR APPLICATION NUMBER: 08/807,151
; PRIOR FILING DATE: 02/27/97
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20030103981A1 556016CD1
; NAME/KEY: unsure
; LOCATION: 235
; OTHER INFORMATION: unknown or other
US-10-235-699-1

Query Match 99.9%; Score 1541; DB 15; Length 283;
Best Local Similarity 100.0%; Pred. No. 2.5e-147;
Matches 283; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLNTSAGNVDIYKLLYHS DACSKRAVSLRCIACGVNINSSRSRIVGSESLPGAMPW 60
Db 1 MKLNTSAGNVDIYKLLYHS DACSKRAVSLRCIACGVNINSSRSRIVGSESLPGAMPW 60
Qy 61 QVSLHVQNVHVCVGGSTITPEWITVAACVCEKPLNPMHTAFAGILROSFMFGAGYQVE 120
Db 61 QVSLHVQNVHVCVGGSTITPEWITVAACVCEKPLNPMHTAFAGILROSFMFGAGYQVE 120
Qy 121 KVISHPNDYDKTKNDIALMKLOKPLTFNDLVKPVCLPMPGMMLOPEOLCMTISGMGATBE 180
Db 121 KVISHPNDYDKTKNDIALMKLOKPLTFNDLVKPVCLPMPGMMLOPEOLCMTISGMGATBE 180
Qy 181 KGTSEVINAARVLLIETORCNSRYVDNLITPAMI CAGFLQGNVDS CGDSGGLVTSK 240
Db 181 KGTSEVINAARVLLIETORCNSRYVDNLITPAMI CAGFLQGNVDS CGDSGGLVTSK 240
Qy 241 NNIMWLIGDTSWGSACAKAYRPGYGNVWVFTDWTYRORADG 283
Db 241 NNIMWLIGDTSWGSACAKAYRPGYGNVWVFTDWTYRORADG 283

RESULT 3

US-09-759-143-934
; Sequence 934, Application US/09759143
; Patent No. US200202248A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jianshun
; APPLICANT: Dillon, David C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jhang, Yugu
; APPLICANT: Henderson, Robert A.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darriek
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasser A.W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C23
; CURRENT APPLICATION NUMBER: US/09/759,143
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 934
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-759-143-934

Query Match 99.8%; Score 1540; DB 9; Length 393;
Best Local Similarity 99.6%; Pred. No. 4.8e-147;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MKLNTSAGNVDIYKLLYHS DACSKRAVSLRCIACGVNINSSRSRIVGSESLPGAMPW 60
Db 111 MKLNTSAGNVDIYKLLYHS DACSKRAVSLRCIACGVNINSSRSRIVGSESLPGAMPW 170
Qy 61 QVSLHVQNVHVCVGGSTITPEWITVAACVCEKPLNPMHTAFAGILROSFMFGAGYQVE 120
Db 171 QVSLHVQNVHVCVGGSTITPEWITVAACVCEKPLNPMHTAFAGILROSFMFGAGYQVE 230
Qy 121 KVISHPNDYDKTKNDIALMKLOKPLTFNDLVKPVCLPMPGMMLOPEOLCMTISGMGATBE 180
Db 231 KVISHPNDYDKTKNDIALMKLOKPLTFNDLVKPVCLPMPGMMLOPEOLCMTISGMGATBE 230
Qy 181 KGTSEVINAARVLLIETORCNSRYVDNLITPAMI CAGFLQGNVDS CGDSGGLVTSK 240
Db 291 KGTSEVINAARVLLIETORCNSRYVDNLITPAMI CAGFLQGNVDS CGDSGGLVTSK 350
Qy 241 NNIMWLIGDTSWGSACAKAYRPGYGNVWVFTDWTYRORADG 283
Db 351 NNIMWLIGDTSWGSACAKAYRPGYGNVWVFTDWTYRORADG 393

RESULT 4

US-09-780-669-934
; Sequence 934, Application US/09780669
; Patent No. US20020051977A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jianshun
; APPLICANT: Dillon, David C.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jhang, Yugu
; APPLICANT: Henderson, Robert A.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.

APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aljun
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Hepler, William
APPLICANT: Hurai, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.427C24
CURRENT FILING DATE: 2001-02-09
NUMBER OF SEQ ID NOS: 943
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 934
LENGTH: 393
TYPE: PRT
ORGANISM: Homo sapiens
US-09-780-669-934

Query Match 99.8%; Score 1540; DB 9; Length 393;
Best Local Similarity 99.6%; Pred. No. 4.8e-147;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKLNTSAGNDVIYKLYHSDACSSKAVSLRCLACGVNLSRSRIVGESSALPGAMPW 60
DB 111 MKLNTSAGNDVIYKLYHSDACSSKAVSLRCLACGVNLSRSRIVGESSALPGAMPW 170
QY 61 QVSLHVQNVHVCSSGSIITPEWITVAHCVCKPLNPMHTAFAGILROSFMFYGAGYQVE 120
DB 171 QVSLHVQNVHVCSSGSIITPEWITVAHCVCKPLNPMHTAFAGILROSFMFYGAGYQVE 230
QY 121 KVISHPNVDKTKNDIALMKLOKPLTFNDLVKPVCLPNPGMWLOPEOLCWSGAGTEE 180
DB 231 KVISHPNVDKTKNDIALMKLOKPLTFNDLVKPVCLPNPGMWLOPEOLCWSGAGTEE 230
QY 181 KGTSEVLANAAVLLIETORCNSRYVDNLTTPMTCAGFLQGVNDSQQSGSGXLYTSK 240
DB 291 KGTSEVLANAAVLLIETORCNSRYVDNLTTPMTCAGFLQGVNDSQQSGSGXLYTSK 350
QY 241 NNIMWLIQDTSWGSCKAKAYRPGYGVNMFVTDWIYRQMRADG 283
DB 351 NNIMWLIQDTSWGSCKAKAYRPGYGVNMFVTDWIYRQMRADG 393

RESULT 5

US-09-822-827-934
Sequence 934, Application US/09822827
Patent No. US20020081680A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.534C1
CURRENT FILING DATE: 2001-03-28
NUMBER OF SEQ ID NOS: 982
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 934
LENGTH: 393
TYPE: PRT
ORGANISM: Homo sapiens
US-09-822-827-934

Query Match 99.8%; Score 1540; DB 9; Length 393;
Best Local Similarity 99.6%; Pred. No. 4.8e-147;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKLNTSAGNDVIYKLYHSDACSSKAVSLRCLACGVNLSRSRIVGESSALPGAMPW 60

DB 111 MKLNTSAGNDVIYKLYHSDACSSKAVSLRCLACGVNLSRSRIVGESSALPGAMPW 170
QY 61 QVSLHVQNVHVCSSGSIITPEWITVAHCVCKPLNPMHTAFAGILROSFMFYGAGYQVE 120
DB 171 QVSLHVQNVHVCSSGSIITPEWITVAHCVCKPLNPMHTAFAGILROSFMFYGAGYQVE 230
QY 121 KVISHPNVDKTKNDIALMKLOKPLTFNDLVKPVCLPNPGMWLOPEOLCWSGAGTEE 180
DB 231 KVISHPNVDKTKNDIALMKLOKPLTFNDLVKPVCLPNPGMWLOPEOLCWSGAGTEE 290
QY 181 KGTSEVLANAAVLLIETORCNSRYVDNLTTPMTCAGFLQGVNDSQQSGSGXLYTSK 240
DB 291 KGTSEVLANAAVLLIETORCNSRYVDNLTTPMTCAGFLQGVNDSQQSGSGXLYTSK 350
QY 241 NNIMWLIQDTSWGSCKAKAYRPGYGVNMFVTDWIYRQMRADG 283
DB 351 NNIMWLIQDTSWGSCKAKAYRPGYGVNMFVTDWIYRQMRADG 393

RESULT 6

US-09-895-793-934
Sequence 934, Application US/09895793
Publication No. US20020192763A1
GENERAL INFORMATION:

APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Matcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuchun
APPLICANT: Kalos, Michael D.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aljun
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Hepler, William T.
APPLICANT: Henderson, Robert A.
APPLICANT: Hurai, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
APPLICANT: Vinals de Basols, Carlota
APPLICANT: Foy, Teresa
APPLICANT: Fanger, Gary R.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.534C2
CURRENT FILING DATE: 2001-06-29
NUMBER OF SEQ ID NOS: 982
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 934
LENGTH: 393
TYPE: PRT
ORGANISM: Homo sapiens
US-09-895-793-934

Query Match 99.8%; Score 1540; DB 10; Length 393;
Best Local Similarity 99.6%; Pred. No. 4.8e-147;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKLNTSAGNDVIYKLYHSDACSSKAVSLRCLACGVNLSRSRIVGESSALPGAMPW 60
DB 111 MKLNTSAGNDVIYKLYHSDACSSKAVSLRCLACGVNLSRSRIVGESSALPGAMPW 170
QY 61 QVSLHVQNVHVCSSGSIITPEWITVAHCVCKPLNPMHTAFAGILROSFMFYGAGYQVE 120
DB 171 QVSLHVQNVHVCSSGSIITPEWITVAHCVCKPLNPMHTAFAGILROSFMFYGAGYQVE 230
QY 121 KVISHPNVDKTKNDIALMKLOKPLTFNDLVKPVCLPNPGMWLOPEOLCWSGAGTEE 180

Db 231 KVTSHPNVDSKTKNDNDIALMKLQKPLTRNDLVKPCVCLPNPGLMLQPEOLCWTISGMGATBE 290
Qy 181 KGTSEVLNAAKVLLIETORCNSRYVYDNLITPAMI CAGFLQGNVDS CGDSGGXLYTSK 240
Db 291 KGTSEVLNAAKVLLIETORCNSRYVYDNLITPAMI CAGFLQGNVDS CGDSGGXLYTSK 350
Qy 241 NNIMWLIGDTSWGSACAKAYRPGYGVGNMVFPTDWTYRQBRADG 283
Db 351 NNIMWLIGDTSWGSACAKAYRPGYGVGNMVFPTDWTYRQBRADG 393

RESULT 7

US-09-895-814-934
Sequence 934, Application US/09895814
Publication No. US20020193296A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yugu
APPLICANT: Kalos, Michael D.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darick
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William T.
APPLICANT: Henderson, Robert A.
APPLICANT: Hurai, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
APPLICANT: Vinals de Baseole, Carlota
APPLICANT: Foy, Teresa
APPLICANT: Panger, Gary R.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.427C26
CURRENT FILING DATE: 2001-06-29
NUMBER OF SEQ ID NOS: 990
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 934
LENGTH: 393
TYPE: PRT
ORGANISM: Homo sapiens
US-09-895-814-934

Query Match 99.8%; Score 1540; DB 10; Length 393;
Best Local Similarity 99.6%; Pred. No. 4.8e-147;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MKLNTSAGNVDIYKKLYHSDACSSRAVSLRCIACGVNINSSRSRIVGSESALPGAWP 60
Db 111 MKLNTSAGNVDIYKKLYHSDACSSRAVSLRCIACGVNINSSRSRIVGSESALPGAWP 170
Qy 61 QVSLHVQNVHVGCGSIIITPEWITVAACHCYEKPLNPMHTAFAAGILRGSFMYGAGYQVE 120
Db 171 QVSLHVQNVHVGCGSIIITPEWITVAACHCYEKPLNPMHTAFAAGILRGSFMYGAGYQVE 230
Qy 121 KVTSHPNVDSKTKNDNDIALMKLQKPLTRNDLVKPCVCLPNPGLMLQPEOLCWTISGMGATBE 180
Db 231 KVTSHPNVDSKTKNDNDIALMKLQKPLTRNDLVKPCVCLPNPGLMLQPEOLCWTISGMGATBE 290
Qy 181 KGTSEVLNAAKVLLIETORCNSRYVYDNLITPAMI CAGFLQGNVDS CGDSGGXLYTSK 240
Db 291 KGTSEVLNAAKVLLIETORCNSRYVYDNLITPAMI CAGFLQGNVDS CGDSGGXLYTSK 350
Qy 241 NNIMWLIGDTSWGSACAKAYRPGYGVGNMVFPTDWTYRQBRADG 283

Db 351 NNIMWLIGDTSWGSACAKAYRPGYGVGNMVFPTDWTYRQBRADG 393

RESULT 8

US-10-012-896-934
Sequence 934, Application US/10012896
Publication No. US20020183251A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yugu
APPLICANT: Kalos, Michael D.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darick
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William T.
APPLICANT: Henderson, Robert A.
APPLICANT: Hurai, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
APPLICANT: Vinals de Baseole, Carlota
APPLICANT: Foy, Teresa
APPLICANT: Panger, Gary R.
APPLICANT: Watanabe, Yoshihiro
APPLICANT: Meagher, Madeleine Joy
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.427C27
CURRENT FILING DATE: 2001-12-10
NUMBER OF SEQ ID NOS: 1011
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 934
LENGTH: 393
TYPE: PRT
ORGANISM: Homo sapiens
US-10-012-896-934

Query Match 99.8%; Score 1540; DB 14; Length 393;
Best Local Similarity 99.6%; Pred. No. 4.8e-147;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MKLNTSAGNVDIYKKLYHSDACSSRAVSLRCIACGVNINSSRSRIVGSESALPGAWP 60
Db 111 MKLNTSAGNVDIYKKLYHSDACSSRAVSLRCIACGVNINSSRSRIVGSESALPGAWP 170
Qy 61 QVSLHVQNVHVGCGSIIITPEWITVAACHCYEKPLNPMHTAFAAGILRGSFMYGAGYQVE 120
Db 171 QVSLHVQNVHVGCGSIIITPEWITVAACHCYEKPLNPMHTAFAAGILRGSFMYGAGYQVE 230
Qy 121 KVTSHPNVDSKTKNDNDIALMKLQKPLTRNDLVKPCVCLPNPGLMLQPEOLCWTISGMGATBE 180
Db 231 KVTSHPNVDSKTKNDNDIALMKLQKPLTRNDLVKPCVCLPNPGLMLQPEOLCWTISGMGATBE 290
Qy 181 KGTSEVLNAAKVLLIETORCNSRYVYDNLITPAMI CAGFLQGNVDS CGDSGGXLYTSK 240
Db 291 KGTSEVLNAAKVLLIETORCNSRYVYDNLITPAMI CAGFLQGNVDS CGDSGGXLYTSK 350
Qy 241 NNIMWLIGDTSWGSACAKAYRPGYGVGNMVFPTDWTYRQBRADG 283
Db 351 NNIMWLIGDTSWGSACAKAYRPGYGVGNMVFPTDWTYRQBRADG 393

RESULT 9

US-09-758-143-932

```
Sequence 932, Application US/09759143
Patent No. US20020022248A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, David C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqi
APPLICANT: Henderson, Robert A.
APPLICANT: Katos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darriek
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.427C23
CURRENT APPLICATION NUMBER: US/09/759,143
CURRENT FILING DATE: 2001-01-12
NUMBER OF SEQ ID NOS: 934
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 932
LENGTH: 492
TYPE: PRT
ORGANISM: Homo sapiens
US-09-759-143-932

Query Match          99.8%; Score 1540; DB 9; Length 492;
Best Local Similarity 99.6%; Pred. No. 6,4e-147;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKLNTSAGNVDIYKCLYHSDACSSKAVVSLRCLACGVNLSRSRQIVGSESLPGAMPW 60
DB 210 MKLNTSAGNVDIYKCLYHSDACSSKAVVSLRCLACGVNLSRSRQIVGSESLPGAMPW 269
QY 61 QVSLHVQNVHVCSSGSIITPEWIVTAHCVKEPLNPMHTAFAGILRQSFMYGAGYQVE 120
DB 270 QVSLHVQNVHVCSSGSIITPEWIVTAHCVKEPLNPMHTAFAGILRQSFMYGAGYQVE 329
QY 121 KVISHPNVDSKTKNDIALMKLOKPLTFNDLVKPVCLPMPGMLOPEOLCWTISGWGATEE 180
DB 330 KVISHPNVDSKTKNDIALMKLOKPLTFNDLVKPVCLPMPGMLOPEOLCWTISGWGATEE 389
QY 181 KGTSEVNAKAVLLIETORCNSRYVDNLITPAMI CAGFLOGNVDS CGDSGGLVTSK 240
DB 390 KGTSEVNAKAVLLIETORCNSRYVDNLITPAMI CAGFLOGNVDS CGDSGGLVTSK 449
QY 241 NNIMWLIGDTSWGSCKAYRPGVYGWVFTDWTYRQWRADG 283
DB 450 NNIMWLIGDTSWGSCKAYRPGVYGWVFTDWTYRQWRADG 492

RESULT 10
US-09-780-669-932
Sequence 932, Application US/09780669
Patent No. US20020051977A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, David C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqi
APPLICANT: Henderson, Robert A.
APPLICANT: Katos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
```

```
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darriek
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William
APPLICANT: Hural, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.427C24
CURRENT APPLICATION NUMBER: US/09/780,669
CURRENT FILING DATE: 2001-02-09
NUMBER OF SEQ ID NOS: 943
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 932
LENGTH: 492
TYPE: PRT
ORGANISM: Homo sapiens
US-09-780-669-932
```

```
Query Match          99.8%; Score 1540; DB 9; Length 492;
Best Local Similarity 99.6%; Pred. No. 6,4e-147;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 MKLNTSAGNVDIYKCLYHSDACSSKAVVSLRCLACGVNLSRSRQIVGSESLPGAMPW 60
DB 210 MKLNTSAGNVDIYKCLYHSDACSSKAVVSLRCLACGVNLSRSRQIVGSESLPGAMPW 269
QY 61 QVSLHVQNVHVCSSGSIITPEWIVTAHCVKEPLNPMHTAFAGILRQSFMYGAGYQVE 120
DB 270 QVSLHVQNVHVCSSGSIITPEWIVTAHCVKEPLNPMHTAFAGILRQSFMYGAGYQVE 329
QY 121 KVISHPNVDSKTKNDIALMKLOKPLTFNDLVKPVCLPMPGMLOPEOLCWTISGWGATEE 180
DB 330 KVISHPNVDSKTKNDIALMKLOKPLTFNDLVKPVCLPMPGMLOPEOLCWTISGWGATEE 389
QY 181 KGTSEVNAKAVLLIETORCNSRYVDNLITPAMI CAGFLOGNVDS CGDSGGLVTSK 240
DB 390 KGTSEVNAKAVLLIETORCNSRYVDNLITPAMI CAGFLOGNVDS CGDSGGLVTSK 449
QY 241 NNIMWLIGDTSWGSCKAYRPGVYGWVFTDWTYRQWRADG 283
DB 450 NNIMWLIGDTSWGSCKAYRPGVYGWVFTDWTYRQWRADG 492
```

```
RESULT 11
US-09-822-827-932
Sequence 932, Application US/09822827
Patent No. US20020081680A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, David C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqi
APPLICANT: Henderson, Robert A.
APPLICANT: Katos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
```

```
Query Match          99.8%; Score 1540; DB 9; Length 492;
Best Local Similarity 99.6%; Pred. No. 6,4e-147;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKLNTSAGNVDIYKCLYHSDACSSKAVVSLRCLACGVNLSRSRQIVGSESLPGAMPW 60
```

```
Db 210 MKLNTSAGNVDIYKGLYHSDACSSKAVVSLRCLACGVNLNSRSRRIYVGSALPGAMPW 269
Qy 61 QVSLHVNVAHVCSSITTPETITVAACVCEKPLNPPHMTAFAGILROSFMFYAGGYVE 120
Db 270 QVSLHVNVAHVCSSITTPETITVAACVCEKPLNPPHMTAFAGILROSFMFYAGGYVE 329
Qy 121 KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLPMPGMLQPEQLCWSGGMATBE 180
Db 330 KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLPMPGMLQPEQLCWSGGMATBE 389
Qy 181 KGKTSVTLNAKVLILFTORCNSRYVDNLITPAMICAGFLQGVNDSGQDSSGGLVTSK 240
Db 390 KGKTSVTLNAKVLILFTORCNSRYVDNLITPAMICAGFLQGVNDSGQDSSGGLVTSK 449
Qy 241 NNIMWLIGDTSWGSCKAYRPGYGVNVPFTDWTYRQMRADG 283
Db 450 NNIMWLIGDTSWGSCKAYRPGYGVNVPFTDWTYRQMRADG 492

RESULT 12
US-09-895-793-932
; Sequence 932, Application US/09895793
; Publication No. US20020192763A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Basols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.534C2
; CURRENT APPLICATION NUMBER: US/09/895, 793
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 932
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-793-932

Query Match 99.8%; Score 1540; DB 10; Length 492;
Best Local Similarity 99.6%; Pred. No. 6,4e-147;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MKLNTSAGNVDIYKGLYHSDACSSKAVVSLRCLACGVNLNSRSRRIYVGSALPGAMPW 60
Db 210 MKLNTSAGNVDIYKGLYHSDACSSKAVVSLRCLACGVNLNSRSRRIYVGSALPGAMPW 269
Qy 61 QVSLHVNVAHVCSSITTPETITVAACVCEKPLNPPHMTAFAGILROSFMFYAGGYVE 120
Db 270 QVSLHVNVAHVCSSITTPETITVAACVCEKPLNPPHMTAFAGILROSFMFYAGGYVE 329
Qy 121 KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLPMPGMLQPEQLCWSGGMATBE 180
Db 330 KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLPMPGMLQPEQLCWSGGMATBE 389
```

```
Db 330 KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLPMPGMLQPEQLCWSGGMATBE 389
Qy 181 KGKTSVTLNAKVLILFTORCNSRYVDNLITPAMICAGFLQGVNDSGQDSSGGLVTSK 240
Db 390 KGKTSVTLNAKVLILFTORCNSRYVDNLITPAMICAGFLQGVNDSGQDSSGGLVTSK 449
Qy 241 NNIMWLIGDTSWGSCKAYRPGYGVNVPFTDWTYRQMRADG 283
Db 450 NNIMWLIGDTSWGSCKAYRPGYGVNVPFTDWTYRQMRADG 492

RESULT 13
US-09-895-814-932
; Sequence 932, Application US/09895814
; Publication No. US20020193296A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Basols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C6
; CURRENT APPLICATION NUMBER: US/09/895, 814
; NUMBER OF SEQ ID NOS: 990
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 932
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-814-932

Query Match 99.8%; Score 1540; DB 10; Length 492;
Best Local Similarity 99.6%; Pred. No. 6,4e-147;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MKLNTSAGNVDIYKGLYHSDACSSKAVVSLRCLACGVNLNSRSRRIYVGSALPGAMPW 60
Db 210 MKLNTSAGNVDIYKGLYHSDACSSKAVVSLRCLACGVNLNSRSRRIYVGSALPGAMPW 269
Qy 61 QVSLHVNVAHVCSSITTPETITVAACVCEKPLNPPHMTAFAGILROSFMFYAGGYVE 120
Db 270 QVSLHVNVAHVCSSITTPETITVAACVCEKPLNPPHMTAFAGILROSFMFYAGGYVE 329
Qy 121 KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLPMPGMLQPEQLCWSGGMATBE 180
Db 330 KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLPMPGMLQPEQLCWSGGMATBE 389
Qy 181 KGKTSVTLNAKVLILFTORCNSRYVDNLITPAMICAGFLQGVNDSGQDSSGGLVTSK 240
Db 390 KGKTSVTLNAKVLILFTORCNSRYVDNLITPAMICAGFLQGVNDSGQDSSGGLVTSK 449
Qy 241 NNIMWLIGDTSWGSCKAYRPGYGVNVPFTDWTYRQMRADG 283
Db 450 NNIMWLIGDTSWGSCKAYRPGYGVNVPFTDWTYRQMRADG 492
```

DB 450 NNIMWLIGDTSWGSCAKAYRPGVYGNVWFTDMITRORADG 492

RESULT 14

US-10-012-896-932

Sequence 932, Application US/10012896

Publication No. US20020183251A1

GENERAL INFORMATION:

APPLICANT: Xu, Jiangchun

APPLICANT: Dillon, David C.

APPLICANT: Mitcham, Jennifer L.

APPLICANT: Harlocker, Susan L.

APPLICANT: Jiang, Yugu

APPLICANT: Kalos, Michael D.

APPLICANT: Retter, Marc W.

APPLICANT: Stolk, John A.

APPLICANT: Day, Craig H.

APPLICANT: Vedvick, Thomas S.

APPLICANT: Carter, Derrick

APPLICANT: Li, Samuel X.

APPLICANT: Wang, Aijun

APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Henderson, Robert A.

APPLICANT: Hurler, John

APPLICANT: McNeill, Patricia D.

APPLICANT: Houghton, Raymond L.

APPLICANT: Vinals de Bassols, Carlot

APPLICANT: Foy, Teresa

APPLICANT: Fanger, Gary R.

APPLICANT: Wantanabe, Yoshihiro

APPLICANT: Meagher, Madeleine Joy

APPLICANT: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

FILE REFERENCE: 210121.427C27

CURRENT APPLICATION NUMBER: US/10/012.896

CURRENT FILING DATE: 2001-12-10

NUMBER OF SEQ ID NOS: 1011

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 932

LENGTH: 492

TYPE: PR

ORGANISM: Homo sapiens

US-10-012-896-932

Query Match 99.8%; Score 1540; DB 14; Length 492;

Best Local Similarity 99.6%; Pred. No. 6.4e-147;

Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 1 MLNTSAGNVDIYKKLYHSDAGSSKAVVSLRCAAGVNLSSRSRIVGGSALPGAMPW 60

DB 210 MLNTSAGNVDIYKKLYHSDAGSSKAVVSLRCAAGVNLSSRSRIVGGSALPGAMPW 269

QY 61 QVSLHVQNVHVCSSIIITPEMIVTAHCVCKPLNPNHMTAFAGILRQSFYAGYQVR 120

DB 270 QVSLHVQNVHVCSSIIITPEMIVTAHCVCKPLNPNHMTAFAGILRQSFYAGYQVR 329

QY 121 KVISHPNYDSKTKNDIALMCKLOKPLTFNDLVKPVCLPNFGMLQPEOLCWSGMAATEE 180

DB 330 KVISHPNYDSKTKNDIALMCKLOKPLTFNDLVKPVCLPNFGMLQPEOLCWSGMAATEE 389

QY 161 KGTSEVLAANAAYLLETORCNSRYVDNLITPAMI CAGLQGNVDSGQDSGGLVTSK 240

DB 390 KGTSEVLAANAAYLLETORCNSRYVDNLITPAMI CAGLQGNVDSGQDSGGLVTSK 449

QY 241 NNIMWLIGDTSWGSCAKAYRPGVYGNVWFTDMITRORADG 283

DB 450 NNIMWLIGDTSWGSCAKAYRPGVYGNVWFTDMITRORADG 492

RESULT 15

US-10-205-823-415

Sequence 415, Application US/10205823

Publication No. US20030108963A1

GENERAL INFORMATION:

APPLICANT: Schlegel, Robert

APPLICANT: Monahan, John B.

APPLICANT: Endege, Wilson O.

APPLICANT: Gannavarapu, Manjula

APPLICANT: Gorbatcheva, Bella

APPLICANT: Hoersch, Sebastian

APPLICANT: Kametkar, Shubhangi

APPLICANT: Monney, Angela M.

APPLICANT: Galt, Karen

APPLICANT: Zhao, Xumei

APPLICANT: Anderson, Dustin

APPLICANT: TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND

TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND

FILE REFERENCE: MRI-044

CURRENT APPLICATION NUMBER: US/10/205.823

CURRENT FILING DATE: 2002-07-25

PRIOR FILING DATE: 2002-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

PRIOR FILING DATE: 2001-07-25

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2003, 18:10:24 ; Search time 18 seconds
(without alignments)
1511.985 Million cell updates/sec

Title: US-09-988-975a-1
Perfect score: 1543
Sequence: 1 MKLNTSAGNVDIYKKLYHSD.....VYGNVWFPTDWIYRQWADG 283

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR 76:*

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------|----------------------|
| 1 | 555.5 | 36.0 | 638 | 1 KQHUP | plasma kallikrein |
| 2 | 541 | 35.1 | 638 | 1 KQRTPL | plasma kallikrein |
| 3 | 536.5 | 34.8 | 1035 | 1 A43090 | enteropeptidase (E |
| 4 | 528.5 | 34.3 | 1019 | 1 A56318 | enteropeptidase (E |
| 5 | 526.5 | 34.1 | 638 | 1 KOMSPL | plasma kallikrein |
| 6 | 520.5 | 33.7 | 417 | 1 K00845 | heparin (EC 3.4.21. |
| 7 | 520 | 33.7 | 625 | 1 KFRHUI | coagulation factor |
| 8 | 509.5 | 33.0 | 855 | 1 JC7731 | membrane-bound arg |
| 9 | 505.5 | 32.8 | 812 | 1 PLMS | plasma (EC 3.4.21 |
| 10 | 503.5 | 32.6 | 416 | 1 S33777 | heparin (EC 3.4.21. |
| 11 | 502.5 | 32.6 | 1034 | 1 A53663 | enteropeptidase (E |
| 12 | 492 | 31.9 | 421 | 1 S11674 | acrosin (EC 3.4.21 |
| 13 | 490 | 31.8 | 1524 | 2 T30337 | polypeptidase - Afri |
| 14 | 481.5 | 31.2 | 415 | 1 A34170 | acrosin (EC 3.4.21 |
| 15 | 480.5 | 31.1 | 436 | 2 JX0172 | acrosin (EC 3.4.21 |
| 16 | 467 | 30.3 | 418 | 2 A37344 | acrosin (EC 3.4.21 |
| 17 | 466.5 | 30.2 | 431 | 2 S47538 | acrosin (EC 3.4.21 |
| 18 | 464.5 | 30.1 | 421 | 2 S25599 | acrosin (EC 3.4.21 |
| 19 | 461.5 | 29.9 | 437 | 2 S18407 | acrosin (EC 3.4.21 |
| 20 | 459 | 29.7 | 1113 | 2 JRB315 | low-density lipoprot |
| 21 | 458.5 | 29.7 | 790 | 1 PLPG | plasma (EC 3.4.21 |
| 22 | 454 | 29.4 | 810 | 2 I46260 | plasma (EC 3.4.21 |
| 23 | 452 | 29.3 | 343 | 1 A57014 | proctasin (EC 3.4. |
| 24 | 449.5 | 29.1 | 810 | 2 B30848 | plasma (EC 3.4.21 |
| 25 | 447.5 | 29.0 | 810 | 2 B61545 | plasma (EC 3.4.21 |
| 26 | 444 | 28.8 | 862 | 1 PLBO | plasma (EC 3.4.21 |
| 27 | 441.5 | 28.6 | 267 | 2 S40006 | trypsin (EC 3.4.21 |
| 28 | 441 | 28.6 | 270 | 2 S56160 | mast cell tryptase |
| 29 | 440 | 28.5 | 275 | 2 S40005 | trypsin (EC 3.4.21 |

| | | | | | |
|----|-------|------|------|----------|--------------------|
| 30 | 439 | 28.5 | 277 | 2 S35340 | trypsin (EC 3.4.21 |
| 31 | 438 | 28.4 | 276 | 2 A38654 | mast cell proteina |
| 32 | 437.5 | 28.4 | 420 | 2 A55283 | acrosin (EC 3.4.21 |
| 33 | 435.5 | 28.2 | 455 | 2 A61545 | plasma (EC 3.4.21 |
| 34 | 435.5 | 28.2 | 810 | 1 PLHU | plasma (EC 3.4.21 |
| 35 | 434.5 | 28.2 | 266 | 2 S54146 | trypsin (EC 3.4.21 |
| 36 | 433.5 | 28.1 | 4548 | 1 S00657 | apoptotrypsin (EC |
| 37 | 432.5 | 28.0 | 274 | 2 JCA171 | trypsin (EC 3.4.2 |
| 38 | 430 | 27.9 | 273 | 2 A47246 | trypsin (EC 3.4.2 |
| 39 | 429 | 27.8 | 274 | 2 S35339 | trypsin (EC 3.4.21 |
| 40 | 428 | 27.7 | 275 | 2 S40007 | trypsin (EC 3.4.21 |
| 41 | 427 | 27.7 | 242 | 2 S49489 | trypsin (EC 3.4.21 |
| 42 | 425 | 27.5 | 263 | 2 A21195 | chymotrypsin (EC 3 |
| 43 | 424 | 27.5 | 263 | 1 KYRFB | chymotrypsin (EC 3 |
| 44 | 424 | 27.5 | 275 | 2 C35863 | trypsin (EC 3.4.2 |
| 45 | 423.5 | 27.4 | 263 | 2 A31299 | chymotrypsin (EC 3 |

ALIGNMENTS

RESULT 1

KQHUP
Plasma kallikrein (EC 3.4.21.34) precursor - human
N:Alternate names: kininogenin; plasma prekallikrein
C:Species: Homo sapiens (man)
C>Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #ext_change 18-Jun-1999
C:Accession: A00921; A37939
R:Chung, D.W.; Fujikawa, K.; McMullen, B.A.; Davie, E.W.
Biochemistry 25, 2410-2417, 1986
A>Title: Human plasma prekallikrein, a zymogen to a serine protease that contains four
A:Reference number: A00921; M01D:86243359; P01D:3521732
A:Accession: A00921
A:Molecule type: mRNA
A:Residues: 1-638 <CHU>
A:Cross-references: GB:M13143; NID:G190262; PIDN:AAA60153.1; PID:G190263
R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.
Biochemistry 30, 2050-2056, 1991
A>Title: Location of the disulfide bonds in human plasma prekallikrein: the presence of
A:Reference number: A37939; M01D:91152016; P01D:1998666
A:Accession: A37939
A:Molecule type: protein
A:Residues: 20-27,40-46,'X',48,'H',50,'X',52-70,'H',75-76,'X',78-80,103-113,131-140,14
;280-283,'X',285,287-291,'X',293-295,314-317,'X',319-320,321-324,'X',329-333,334-339,'
523;538-551,562,'X',564-567,573,'X',575-576,578-583,'X',585,592-604 <MCM>
C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex
C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a
C:Comment: The enzyme cleaves Lys-Arg and Arg-Ser bonds. It activates, in a reciprocal
inogen and may also play a role in the renin-angiotensin system by converting prorenin
C:Genetics:
A:Gene: GDB:TLK3
A:Cross-references: GDB:127575; OMIM:229000
A:Map position: 4q35-4q35
C:Superfamily: coagulation factor XI; trypsin homology
C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; inf
F:1-10/Domain: signal sequence #status predicted <SIG>
F:20-338/Product: plasma kallikrein #status predicted <MAT>
F:20-390/Domain: plasma kallikrein heavy chain #status predicted <HCH>
F:20-109/Domain: apple repeat <AP1>
F:110-199/Domain: apple repeat <AP2>
F:200-289/Domain: apple repeat <AP3>
F:291-380/Domain: apple repeat <AP4>
F:391-638/Domain: plasma kallikrein light chain #status predicted <LCH>
F:391-623/Domain: trypsin homology <TRY>
F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-322,328,3
F:127,308,336,433,494/Binding site: carboxylate (Asn) (covalent) #status experimental
F:318-347,340-345/Disulfide bonds: #status predicted
F:390-391/Cleavage site: Arg-Ile (coagulation factor XIIa) #status predicted
F:434,483,578/Active site: Arg, Asp, Ser #status predicted

Query Match
Best Local Similarity 36.0%; Score 555.5; DB 1; Length 638;
Pred. 41.8%; Pred. No. 2.4e-42;

[illegible]

RESULT 2

A:Accession: A33180
A:Molecule type: DNA
A:Residues: 1-638 <BEA>
A:Cross-references: GB:J05315
A>Note: the authors translated the codon GAG for residue 81 as Glu
R:Setdah, N.G.; Ladsenheim, R.; Mbikey, M.; Hamelin, J.; Lutalla, G.; Rougeon, F.; Lazunee, F.
DNA 8, 563-574, 1989
A>Title: The cDNA structure of rat plasma kallikrein.
A:Reference number: A33320; MUID:90091743; PMID:2598771
A:Accession: A33320
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-638 <SBI>
A:Cross-references: GB:N30282; NID:g205010; PIDN:AAA1463.1; PID:g205011
A>Note: part of this sequence, including the amino ends of both the heavy and light chains
R:Papain, J.; Benjamin, S.; Sawyer, N.; Lazure, C.; Chretien, M.; Setdah, N.G.
Biochim. Biophys. Acta 999, 103-110, 1989
A>Title: Rat plasma kallikrein purification, NH(2)-terminal sequencing and development
A:Reference number: S06851; MUID:90089457; PMID:2597701
A:Accession: S06851
A:Molecule type: protein
A:Residues: 20-45/391-413 <PAO>
R:Setdah, N.G.; Ladsenheim, R.; Mbikey, M.; Hamelin, J.; Lutalla, G.; Rougeon, R.; Lazunee, F.
DNA Cell Biol. 8, 563-574, 1989
A>Title: The cDNA structure of rat plasma kallikrein.
A:Reference number: I53041
A:Accession: I53041
A>Status: translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-638 <RES>
A:Cross-references: GB:M58590; NID:g206721; PIDN:AAA2069.1; PID:g206722
A:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex with factor XIa,
C:Comment: The zymogen is activated by factor XIa, which cleaves the molecule into a 1:
are linked by one or more disulfide bonds.
C:Genetics:
A:Gene: PK
C:Superfamily: coagulation factor XI, trypsin homology
C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla
F:1-19/Domains: signal sequence #status predicted <SI>

F.220-390/Product plasma kallikrein heavy chain #status experimental <MAT1>
F.220-109/Domain: apple repeat <AP1>
F.110-199/Domain: apple repeat <AP2>
F.200-289/Domain: apple repeat <AP3>
F.291-380/Domain: apple repeat <AP4>
F.391-638/Product: plasma kallikrein light chain #status experimental <MAT2>
F.391-632/Domain: trypsin homology <TRY>
F.121-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,3
F.127,215,308,453,459,494/Binding site: carboxydrate (Asn) (covalent) #status predicece
F.396/Binding site: carboxydrate (Asn) (covalent) #status experimental
F.434,463,578/Active site: His, Asp, Ser #status predicted

Query Match

Best Local Similarity 38.2%; Pred.No. 4.9e-41;
Matches 109; Conservative 54; Mismatches 112; Indels 10; Gaps 5;

QY 1 MC LNTSAGNVDIYKCLYHSDACSSKAVVSLRCIACGVNLTNSSRQRI VGGESALPGAMPW 60

Db 349 LRISTDGSPTRTTYEAQGSSSGYSLRLCKRVSSSDCTTKIN---ARIVGGTNNSSLGEMPM 404

61 QVSLHVQNV---HVCGSLTPBWLVAACVKEKLNPFHWIAFAGLRQSFMEYGAGY 11

ZOOLOGICAL MUSEUM OF THE UNIVERSITY OF TORONTO

[illegible]

1

[illegible]

237 VMSKNJWLTIGNTSWSGCAKAYBPAVYGVNVMETDWTYBOMPA 281

Db
583 VCKHSGRWLVGITSWEGGCARKEOPGVTTKVAEYIDNILEKIOS 6227

RESULT 2

enteropeptidase (EC 3.4.21.9) precursor [validated] - bovine

C/Species: Bos primigenius taurus (cattle)
C/Date: 10-Sep-1998 #sequence revision 10-Sep-1999 #text change 28-Apr-2002

C;ACCESSION: A43090; A488/4; A61436
R;Kitamoto, Y.; Yuan, X.; Wu, O.; McCourt, D.W.; Sadler, J.E.

A; Title: Enterokinase, the initiator of intestinal digestion, is a mosaic p

A/Accession: A43090

A;Molecule type: mRNA

A; Cross-references: GB:U09859; NID:g746410; PIDN:AAB40026.1; PID:g746411

R.; Lavallie, B.R.; Renemtulla, A.; Racine, L.A.; DiBlasio, E.A.; Ferenz, C.; J. Biol. Chem. 268: 23311-23317, 1993

A;Reference number: A48874; MUID:94043122; PMID:8226855

A;Molecule type: mRNA

A/Cross-references: GB:L19663; NID:g416131; PIDN:AAA16035.1; PID:g416132

R;Light, A.; Janska, H.
T. Dactylina Chem 10 475-480 1991

A;Title: The amino-terminal sequence of the catalytic subunit of bovine ent
A;Reference number: A61436: MTTD:92189715: PMTD:1799406

A;Molecule type: protein
A;Accession: A61438

C;Comment: The mechanism of association with the membrane of the intestinal

C;Comment: Conversion from membrane-bound to soluble forms may involve furt

C/Complex: mature enteropeptidase is variously reported to contain two (heavy and light) hide linked
C/Function:
A/Description: cleaves propeptide from trypsinogen to produce active trypsin
A/Pathway: intestinal digestive hydrolase cascade
C/Superfamily: enteropeptidase; Clr/Cis repeat homology; LDL receptor ligand-binding rep
C/Keywords: glycoprotein; hydrolase; intestine; serine proteinase; transmembrane protein
F:22-38/Domain: transmembrane #status predicted <TM>
F:52-117/Product: enteropeptidase mini chain #status predicted <MCH>
F:118-800/Product: enteropeptidase heavy chain #status predicted <HCH>
F:199-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:358-520/Domain: MAM homology <MAM>
F:659-693/Domain: Clr/Cis repeat homology <Clr>
F:694-799/Domain: LDL receptor ligand-binding repeat homology #status atypical <SRC
F:801-1035/Product: enteropeptidase light chain #status predicted <LCH>
F:801-1030/Domain: trypsin homology <TRY>
F:116-147,170,194,233,263,264,404,456,486,519,550,646,698,722,741,762,864,903,965/Bindin
F:188-912,826-842,926-993,957-972,983-1011/Disulfide bonds: #status predicted
F:841,892,987/Active site: His, Asp, Ser #status predicted

Query Match 34.8%; Score 536.5; DB 1; Length 1035;
Best Local Similarity 39.2%; Pred. No. 2,3e-40;
Matches 111; Conservative 53; Mismatches 104; Indels 15; Gaps 6;

QY 1 MKLNTSAGNVDIYKLYHSDACSSKAVSLRC--IAGVNLSSRS--RIVGESALPGA 57
Db VLNLTAPNSLI---LTPEOQLDLSLILQCNKSGKGLVQVSPKIVGSDSREGA 811
58 WPMQVSLHVNHHVCGSITTEPMITVAHCVKPLNPNMHTAFAGILRQSFMPYAGY 117
Db WPMVVALYDDQVCGASLVSRDMLVSAHCVGRMBPSKMAVGLHNSNL---TSP 868
QY 118 QVE-----KVISHPNYDSTKNNNDIALMLOKPLTNDLVKPVCLPNPGMTLOPBDLCWI 172
Db QIETLLIQDVIINPHNKRKKNNDIAMHLEKVNVTDYIOPCLPEBNQVFPFGRICSI 928
QY 173 SGMGATTEKGTSEVLANAKVLLIETORCNSRYVDNLTTPAMICGFLQGNVDSQSGS 232
Db AGMGALLIYQGSTADVQLQBADVPLLSNEKQQQCMPEYN--ITENNVCAGYERAGVDSQSGS 987
QY 233 GGLVTSKNNIMWLLIGDTSMGSGCAKAVRPGYGVNMTVFTDWI 275
Db 988 GSPLMQENRMLAGVTSFGYCALPNRPGYARVRFTETWI 1030

RESULT 4
A56318
enteropeptidase (EC 3.4.21.9) precursor [validated] - human
N/Alternate names: enterokinase
C/Species: Homo sapiens (man)
C/Date: 19-May-1995 #sequence__revision 09-Aug-1996 #text_change 28-Apr-2003
C/Accession: A56318; B43090
R/Kitamoto, Y.; Velle, R.A.; Donis-Keller, H.; Sadler, J.E.
Biochemistry 34, 4562-4568, 1995
A/Title: cDNA sequence and chromosomal localization of human enterokinase, the proteolyt
A/Reference number: A56318; MUID:95234679; PMID:7718557
A/Accession: A56318
A/Molecule type: mRNA
A/Residues: 1-1019 <KIT>
A/Cross-references: GB:U09860; NID:9746412; PIDN:AA650138.1; PID:9746413
R/Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.
Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994
A/Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease compo
A/Reference number: A43090; MUID:94329561; PMID:8052624
A/Accession: B43090
A/Status: nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 749-1019 <KIT>
A/Cross-references: GB:U09860
C/Comment: The mechanism of association with the membrane of the intestinal brush border
C/Genetics: (noted below) or with amino-terminal myristoylation of the heavy chain.

A/Gene: GDB:PRS57
A/Cross-references: GDB:384083; OMIM:226200
A/Map position: 21q21-21q21
C/Complex: Mature enteropeptidase is variously reported to contain two (heavy and lig
ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms inv
duces.
C/Function:
A/Description: cleaves activation peptide from trypsinogen to produce active trypsin
A/Pathway: intestinal digestive hydrolase cascade
C/Superfamily: enteropeptidase; Clr/Cis repeat homology; LDL receptor ligand-binding
C/Keywords: glycoprotein; hydrolase; intestine; serine proteinase; transmembrane protein; zymo
F:1-784/Product: enteropeptidase heavy chain #status predicted <HCH>
F:22-38/Domain: transmembrane #status predicted <TM>
F:184-221/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:342-504/Domain: MAM homology <MAM>
F:526-631/Domain: Clr/Cis repeat homology <Clr>
F:643-677/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:678-783/Domain: scavenger receptor cysteine-rich domain homology #status atypical <
F:785-1019/Product: enteropeptidase light chain #status predicted <LCH>
F:785-1014/Domain: trypsin homology <TRY>
F:116,147,179,328,335,388,440,470,503,534,630,682,706,725,848,887,909,949/Binding sit
F:172-896,810-826,910-977,941-956,967-995/Disulfide bonds: #status predicted
F:925,876,971/Active site: His, Asp, Ser #status predicted

Query Match 34.3%; Score 528.5; DB 1; Length 1019;
Best Local Similarity 38.6%; Pred. No. 1.2e-39;
Matches 108; Conservative 56; Mismatches 107; Indels 9; Gaps 5;

QY 1 MKLNTSAGNVDIYKLYHSDACSSKAVSLRC--IAGVNLSSRS--QSRIVGESALPGA 57
Db VLNLTAP---DGLILTPEOQLDLSLILQCNKSGKGLAODITPKIVGSSNAKEGA 795
QY 58 WPMQVSLHVNHHVCGSITTEPMITVAHCVKPLNPNMHTAFAGILRQSFMPF--YGA 115
Db WPMVVALYGGRLCGASLVSSDMLVSAHCVGRMLBPSKMAIIGLHNSNLTPQTV 855
QY 796 WPMVVALYGGRLCGASLVSSDMLVSAHCVGRMLBPSKMAIIGLHNSNLTPQTV 855
Db 796 WPMVVALYGGRLCGASLVSSDMLVSAHCVGRMLBPSKMAIIGLHNSNLTPQTV 855
QY 116 QYQVEKVIHNPYDSTKNNNDIALMLOKPLTNDLVKPVCLPNPGMTLOPBDLCWISG 175
Db PRILDEIVINPHNKRKKNNDIAMHLEKVNVTDYIOPCLPEBNQVFPFGRICSI 915
QY 856 GATEBKGTSEVLANAKVLLIETORCNSRYVDNLTTPAMICGFLQGNVDSQSGS 235
Db 176 GATEBKGTSEVLANAKVLLIETORCNSRYVDNLTTPAMICGFLQGNVDSQSGS 235
QY 916 GTVVGQGTANTLQBADVPLLSNERCQQQCMPEYN--ITENNVCAGYERAGVDSQSGS 974
Db 916 GTVVGQGTANTLQBADVPLLSNERCQQQCMPEYN--ITENNVCAGYERAGVDSQSGS 974
QY 236 LVTSKNNIMWLLIGDTSMGSGCAKAVRPGYGVNMTVFTDWI 275
Db 975 LMCQENRMLAGVTSFGYCALPNRPGYARVRFTETWI 1014

RESULT 5
K0MSPL
Plasma kallikrein (EC 3.4.21.34) precursor - mouse
C/Species: Mus musculus (house mouse)
C/Date: 30-Sep-1992 #sequence__revision 30-Sep-1992 #text_change 18-Jun-1999
C/Accession: A36557
R/Seldah, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaudien, G.; Brachpapa, L.; Rocher
DNA Cell Biol. 9, 737-748, 1990
A/Title: Mouse plasma kallikrein: cDNA structure, enzyme characterization, and compari
A/Reference number: A36557; MUID:91090844; PMID:2264928
A/Accession: A36557
A/Molecule type: mRNA
A/Residues: 1-638 <SEI>
A/Cross-references: GB:M58589; NID:9200358; PIDN:AA63393.1; PID:9200359
A/Note: Part of this sequence, including the amino ends of both the heavy and light ct
C/Comment: This protein, synthesized in the liver, circulates as a noncovalent complex
C/Complex: The zymogen is activated by factor XII, which cleaves the molecule into a
cleaved by one or more disulfide bonds.
C/Superfamily: coagulation factor XI; trypsin homology
C/Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; inf
F:1-19/Domain: signal sequence #status predicted <Sig>
F:20-39/Product: plasma kallikrein heavy chain #status experimental <HCH>
F:20-109/Domain: apple repeat <AP1>
F:110-199/Domain: apple repeat <AP2>

F:200-289/Domain: apple repeat <AP3>
 F:291-380/Domain: apple repeat <AP4>
 F:391-638/Product: plasma kallikrein light chain #status experimental <LCH>
 F:391-624/Domain: trypsin homology <TRY>
 F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322
 F:127,215,308,396,494/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:433,483,578/Active site: His, Asp, Ser #status predicted

Query Match 34.1%; Score 526.5; DB 1; Length 638;
 Best Local Similarity 39.1%; Pred. No. 1e-39;
 Matches 106; Conservative 46; Mismatches 96; Indels 23; Gaps 5;

QY 15 KLYHSDASSKAVSLRCLACGVNLSRSQSRIVGESALPGAMPQVSLAHQNV---HY 71
 DB 376 KLVDSPDCTTK-----NARIVGGTNSLSLGPMPQVSLQVLSQTHL 418
 QY 72 CGGSITTEMTVTAACHCKEPLNNPMTAFAGILQSFMPFYGA-GYQVEKVISHPNYS 130
 DB 419 CGGSITIGKQWVLTAAHCPD-GIPYDPVWRIYGGILSLSEITETPSRIKELIHQRYKY 477
 QY 131 KTKNDIALMKLOKPLTFNDLVKPVCLPMPGMLOEOLCWISGAGTEBKGTSEVLTNA 190
 DB 478 SEGNDIALIKLOPLNTYTFEOKFICLPKADTNTITTCWYTGWGTKEQSTQNILOK 537
 QY 191 AKVLLIEFQRCNSRYVDLITPAMICAGFLQGNVSCGDSGKLVTSKNNITWLIQDT 250
 DB 538 ATIPVPEECCCKY-RDVIVKMKICAGYKCEGTACKGDSGGLVCKHSGRWQLVGIT 596
 QY 251 SWGSCAKAYRPGVYGVNVVFTDVIYQKRA 281
 DB 597 SWGSCGKRDQPGVITKVSRYMDIIEKQOS 627

RESULT 6

hepsin (BC 3.4.21.-) - human
 S00845
 C:Species: Homo sapiens (man)
 C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #ext_change 18-Jun-1999
 C:Accession: S00845
 R:Leytus, S.P.; Jacob, K.R.; Hagen, F.S.; Kurachi, K.; Davie, E.W.
 Biochemistry 27, 1067-1074, 1988
 A:Title: A novel trypsin-like serine protease (hepsin) with a putative transmembrane dom
 A:Reference number: S00845; PMID:88209431; PMID:2835076
 A:Accession: S00845
 A:Molecule type: mRNA
 A:Residues: 1-417 <LEYS>
 C:Genetics:
 A:Gene: GDB:HPN; TMPRSS1; hepsin
 A:Cross-references: GDB:135685; OMIM:142440
 A:Map position: 19q11-19q13.2
 C:Superfamily: hepsin; trypsin homology
 C:Keywords: hydrolase; liver; serine proteinase; transmembrane protein
 F:23-45/Domain: transmembrane #status predicted <TM>
 F:163-400/Domain: trypsin homology <TRY>
 F:188-204,281-359,322-338,349-381/Disulfide bonds: #status predicted
 F:203,257,353/Active site: His, Asp, Ser #status predicted

Query Match 33.7%; Score 520.5; DB 1; Length 417;
 Best Local Similarity 38.7%; Pred. No. 2.1e-39;
 Matches 104; Conservative 43; Mismatches 109; Indels 13; Gaps 4;
 QY 22 CSSKAVSVSLRCLACGVNLSRSQSRIVGESALPGAMPQVSLAHQNVHVCSSITTEW 81
 DB 140 CRRGRFLAICODCG--RRKLTPDRIVGSDTSLGMPQVSLRYDGAHLCCGSLSGDW 197
 QY 82 IYTAHCVCKPLNNPMTAFAGILQSFMPFYGA-GYQVEKVISHPNY-----DSKTKN 135
 DB 198 VLTAAHCPERRRNVSRMVFACAVQAQAP-RKLQCGVAVVHGGELPRDENSENS 256
 QY 136 DIALMKLOKPLTFNDLVKPVCLPMPGMLOEOLCWISGAGTEBKGTSEVLTNAKVL 195
 DB 257 DIALVHLSPLPTEITIGVCLPAAQALVDGKICVTYTGKNTGYTGQAGVLOEARKVPI 316

QY 196 IETQRCNSRYVDLITPAMICAGFLQGNVSCGDSGKLV-----TSKNNITWLIQDTS 251
 DB 317 ISNDVCGADFYGNQIKPEMPCAGIPESGIDACQDSSGSPFVCEBSISHPRLCGIVS 376
 QY 252 WSGCAKAYRPGVYGVNVVFTDVIYQKRA 280
 DB 377 WGTGCLAKQKPGVITKVSDFREMFQAIK 405

RESULT 7

KFNU1
 N:Alternate names: antihemophilic factor C; plasma thromboplastin antecedent
 C:Species: Homo sapiens (man)
 C:Date: 13-Aug-1986 #sequence_revision 26-May-1994 #ext_change 08-Dec-2000
 C:Accession: A27431; A00920; A37940
 R:Asakai, R.; Davie, E.W.; Chung, D.W.
 Biochemistry 26, 7221-7228, 1987
 A:Title: Organization of the gene for human factor XI.
 A:Reference number: A27431; PMID:88107663; PMID:2827746
 A:Accession: A27431
 A:Molecule type: DNA
 A:Residues: 1-625 <ASA>
 A:Cross-references: GB:M18295
 A:Note: The sequence shown follows the authors' translation
 R:Fujikawa, K.; Chung, D.W.; Hendrickson, L.E.; Davie, E.W.
 Biochemistry 25, 2417-2424, 1986
 A:Title: Amino acid sequence of human factor XI, a blood coagulation factor with four
 A:Reference number: A00920; PMID:86243360; PMID:3636155
 A:Accession: A00920
 A:Molecule type: mRNA
 A:Residues: 1-625 <FUJ>
 A:Cross-references: GB:M13142; NID:9182832; PIDN:AAA52487.1; PID:9182833
 R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.
 Biochemistry 30, 2056-2060, 1991
 A:Title: Location of the disulfide bonds in human coagulation factor XI: the presence
 A:Reference number: A37940; PMID:91152017; PMID:1998667
 A:Accession: A37940
 A:Molecule type: protein
 A:Residues: 28-33;35-49,'X',51-55,'X',57-63;70-75,'X',77-79;107-109,'X',111-112;132-13
 C:Comment: The proenzyme consists of two identical chains linked by one or more disulf
 he active site, and a heavy chain, which associates with high molecular weight (HMW) k

C:Genetics:
 A:Gene: GDB:F11
 A:Cross-references: GDB:119891; OMIM:264900
 A:Map position: 4q35-4q35
 A:Introns: 19/1; 73/2; 109/1; 162/2; 199/1; 252/2; 289/1; 343/2; 379/1; 435/2; 494/1;
 C:Function:
 A:Description: catalyzes the proteolytic activation of coagulation factor IX
 A:Pathway: blood coagulation intrinsic pathway
 C:Superfamily: coagulation factor XI; trypsin homology
 C:Keywords: blood coagulation; duplication; glycoprotein; hemophilia C; homodimer; hyd
 F:1-18/Domain: signal sequence #status predicted <SIG>
 F:19-387/Product: coagulation factor Xla heavy chain #status experimental <HCH>
 F:19-108/Domain: apple repeat <AP1>
 F:109-198/Domain: apple repeat <AP2>
 F:199-288/Domain: apple repeat <AP3>
 F:290-379/Domain: apple repeat <AP4>
 F:388-625/Product: coagulation factor Xla light chain #status experimental <LCH>
 F:388-618/Domain: trypsin homology <TRY>
 F:20-103,514-581,571-599/Disulfide bonds: #status predicted
 F:29/Disulfide bonds: interchain #status experimental
 F:46-76,50-56,110-193,136-165,140-146,200-283,226-255,230-236,291-374,317-346,321-327,
 F:90,126,353,450/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:339/Disulfide bonds: interchain #status predicted
 F:387-388/Cleavage site: Arg-11e (coagulation factor Xla) #status experimental
 F:431,480,575/Active site: His, Asp, Ser #status predicted
 F:491/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 33.7%; Score 520; DB 1; Length 625;
 Best Local Similarity 41.8%; Pred. No. 3.9e-39;

Db 549 RKLX--DYCDIPLCASASSFECGKQVPEPKKCPGRVGGCVANPHSWPMQISLRTFTGQ 606
 Qy 70 HVGCGSIIITPRTITAAHCVEKPLANNPMTAIPAGILROSPFPGAGVOE-----KITS 124
 Db 607 HFCGGTILAPEMVITAAHLEKRS--SRPFYVITLGAHRE---YLRGLDVEISVAKIL 661
 Qy 125 HPNVDSKTKQNDIALMKLOKPLTFNDLVKPVCLPFGMMLQPEQLCWISMGATE---EK 181
 Db 662 EPN-----NRDILALKSRPATITTDKVIPLCLSPNNMVAADRTICVITMGEGTGTGA 715
 Qy 182 GKTSEVINAAYKLLIETQRCNSRYVDNLITPAMICAGFLQGVNDSGCGSGXLYTSKN 241
 Db 716 GRLEK-----AQLPIENKVCGRVEYLANNRVSTELCAGQLAGVDSGCGSGSLVCEK 771
 Qy 242 NIMWLIGDTSWGSGCAKAYRPGYGVNMFPTDITRQNR 280
 Db 772 DKYILQGVTSWGLGCAIRPKDGVYVRSRPFVDIRERNR 810

RESULT 10

33377 hepsin (EC 3.4.21.-) - rat
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Jun-1999
 C:Accession: S33777; S32013
 R:Barley, D.; Raymond, F.; Nick, H.
 Biochim. Biophys. Acta 1173, 350-352, 1993
 A:Title: Cloning and sequence analysis of rat hepsin, a cell surface serine proteinase.
 A:Reference number: S33777; MUID:93305733; PMID:8318546
 A:Accession: S33777
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-416 <FAR>
 A:Cross-references: EMBL:X70900; NID:957928; PIDN:CA50256.1; PID:957929
 C:Superfamily: hepsin; trypsin homology
 C:Keywords: hydrolyase; liver; serine proteinase; transmembrane protein
 F:22-44/Domain: transmembrane #status predicted <TM>
 F:162-199/Domain: trypsin homology <TRY>
 F:187-203,290-358,321-337,348-380/Disulfide bonds: #status predicted
 F:202,256,352/Active site: His, Asp, Ser #status predicted

Query Match 32.6%; Score 503.5; DB 1; Length 416;
 Best Local Similarity 37.9%; Pred. No. 7.5e-38;
 Matches 102; Conservative 45; Mismatches 109; Indels 13; Gaps 4;
 Qy 22 CSSRAVSLRACIAGVNLSSRSRIVGESALPGAMFVQSLAVQNVHVGSGSIITPEW 81
 Db 139 CPGREFILATCODCG--RRKLPIVDRIVGGQDSLSGRMFQVSLRYDGTILCGSILSGDW 196
 Qy 82 IVTAAHCVEKPLANNPMTAIPAGILROSPFPGAGVOEYKISHPNY-----DSKTKN 135
 Db 197 VLTAAHCPEERNRVLSSRRVPAAGAVARTSP-HAVQLGVQAVYHGGYLPFRDPTIDENSN 255
 Qy 136 DIAMKLOKPLTFNDLVKPVCLPFGMMLQPEQLCWISMGATEKGTSEVINAAYKLL 195
 Db 256 DIALVHLSSTPLREYIQPVCLPAAQALVDSKVLTVGMGTQRYGGQAAVLLQGRVPI 315
 Qy 196 IETQRCNSRYVDNLITPAMICAGFLQGVNDSGCGSGXLY---TSKNIMWLIGDTS 251
 Db 316 ISNEVCSNPDYFGNQIKKMFCAGYRPEGIDACGDSGHPFCEDRISGTSNMRICGLVS 375
 Qy 252 WSGGCAKAYRPGYGVNMFPTDITRQNR 280
 Db 376 WGTGCAIARKEGYVTKVIDFRFWIQAIR 404

RESULT 11

AS3663 enteropeptidase (EC 3.4.21.9) precursor [validated] - pig
 N:Alternate names: enterokinase
 C:Species: Sus scrofa domestica (domestic pig)
 C>Date: 07-Oct-1994 #sequence_revision 09-Aug-1996 #text_change 28-Apr-2003
 C:Accession: A53663

R:Matsumura, M.; Ichinose, M.; Yahagi, N.; Kakei, N.; Tsukada, S.; Miki, K.; Kurokawa
 J. Biol. Chem. 269, 19976-19982, 1994
 A:Title: Structural characterization of porcine enteropeptidase.
 A:Reference number: A53663; MUID:94327548; PMID:8051081
 A:Accession: A53663
 A:Molecule type: mRNA
 A:Residues: 1-1034 <MAT>

A:Cross-references: GB:D30799; NID:9505122; PIDN:BA06459.1; PID:9505123
 A>Note: parts of this sequence, including the amino ends of three chains isolated from
 C:Comment: The mechanism of association with the membrane of the intestinal brush bord
 (located below) or with amino-terminal myristoylation of the heavy chain.
 C:Complex: Mature enteropeptidase is variously reported to contain two (heavy and light
 ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms invo
 C:Function:
 A:Description: cleaves activation peptide from trypsinogen to produce active trypsin
 A:Pathway: intestinal digestive hydrolyase cascade
 C:Superfamily: enteropeptidase; Ctr/Cls repeat homology; LDL receptor ligand-binding r
 C:Keywords: glycoprotein; hydrolyase; serine proteinase; transmembrane protein; zymogen
 F:22-38/Domain: transmembrane #status predicted <TM>
 F:52-117/Product: enteropeptidase mini chain #status predicted <MCH>
 F:118-799/Product: enteropeptidase heavy chain #status predicted <HCH>
 F:119-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>
 F:357-519/Domain: MAM homology <MAM>
 F:541-646/Domain: Ctr/Cls repeat homology <CTR>
 F:658-692/Domain: LDL receptor ligand-binding repeat homology <LDL2>
 F:693-798/Domain: scavenger receptor cysteine-rich domain homology #status atypical <S
 F:800-1034/Product: enteropeptidase light chain #status predicted <LCH>
 F:800-1029/Domain: trypsin homology <TRY>
 F:116,147,170,194,283,343,350,403,455,485,518,549,645,697,701,721,740,761,804,863,902,
 F:787-911,825-841,925-992,956-971,982-1010/Disulfide bonds: #status predicted
 F:840,891,986/Active site: His, Asp, Ser #status predicted

Query Match 32.6%; Score 502.5; DB 1; Length 1034;
 Best Local Similarity 36.4%; Pred. No. 2.7e-37;
 Matches 103; Conservative 60; Mismatches 105; Indels 15; Gaps 6;
 Qy 1 MKLNTSAGNDVIKKLYHSDACSSKAVVSLNC--IACGVNLSSSQS--RIYGESALRKA 57
 Db 754 VKLNTAPNGSLI---LTASBECFFEDSLTLQCNHKSCKGKQVQAQSVPTVGGNDSREGA 810
 Qy 58 WPMQVSLHVNQVHVGSGSIITPEMIVTAACVEKPLANNPMTAIPAGILROSPFPGAGY 117
 Db 811 WPMVVALYVNOQLCGASLVGRDMLVSAHCVYGRNLEPSKRALGLHMTSNL---TSP 867
 Qy 118 Q-----VEKVISHPYDSKTKQNDIALMKLOKPLTFNDLVKPVCLPFGMMLQPEQLCWI 172
 Db 868 QIVRLIDRIVINPHYNNRRROSDIAMWHLEFKVYVTDYIPICLPEENQVPPERICSI 927
 Qy 173 SGWATEKGTSEVINAAYKLLIETQRCNSRYVDNLITPAMICAGFLQGVNDSGCGDS 232
 Db 928 AGMGKVYIQGSPADILQADVPLLSNEKCOQOMPEYN--ITENMCAGYBEGGIDSCGDS 986
 Qy 223 GGLVTSKNNIMWLIGDTSWGSGCAKAYRPGYGVNMFPTDITRQNR 275
 Db 987 GGLPMLCEENRMLLAGVTSFQYQALPNKPGYAVARVPKFTENI 1029

RESULT 12

S11674 acrosin (EC 3.4.21.10) precursor - human
 C:Species: Homo sapiens (man)
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
 C:Accession: S11674; S23499; S12063; A61022; S03330
 R:Keime, S.; Adham, I.M.; Engel, W.
 Eur. J. Biochem. 190, 195-200, 1990
 A:Title: Nucleotide sequence and exon-intron organization of the human proacrosin gene
 A:Reference number: S11674; MUID:90306003; PMID:2114285
 A:Accession: S11674
 A:Molecule type: DNA
 A:Residues: 1-421 <KEI>
 A:Cross-references: EMBL:X54017; NID:935582; PIDN:CA437964.1; PID:91216165
 A>Note: the authors translated the codon AGG for residue 64 as Thr and CTG for residue
 R:Varquez-Levin, M.H.; Reventoe, J.; Gordon, J.W.

Eur. J. Biochem. 207, 23-26, 1992
 A>Title: Molecular cloning, sequencing and restriction mapping of the genomic sequence
 A:Reference number: S23499; MUID:92331659; PMID:1628652
 A:Accession: S23499
 A>Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-421 <V42>
 A:Cross-references: EMBL:M77378
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1992.
 R:Kame, S.
 Submitted to the EMBL Data Library, December 1989
 A:Reference number: S12063
 A:Accession: S12063
 A:Molecule type: DNA
 A:Residues: 1-225, 'R', 227-421 <KR12>
 A:Cross-references: EMBL:X54017
 R:Adham, I.M.; Klemm, U.; Maier, W.M.; Engel, W.
 Hum. Genet. 84, 125-128, 1990
 A>Title: Molecular cloning of human proacrosin cDNA.
 A:Reference number: A61022; MUID:90128988; PMID:2298447
 A:Accession: A61022
 A>Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-63, 'T', 65-225, 'V', 227-267, 'R', 269-421 <ADR>
 R:Baba, T.; Watanabe, K.; Kashiwabara, S.I.; Arai, Y.
 FEBS Lett. 244, 296-300, 1989
 A>Title: Primary structure of human proacrosin deduced from its cDNA sequence.
 A:Reference number: S03330; MUID:89153568; PMID:2493394
 A:Accession: S03330
 A:Molecule type: mRNA
 A:Residues: 1-63, 'T', 65-119, 'V', 121-165, 'L', 167-267, 'R', 269-344, 'R', 346-421 <BAB>
 A:Cross-references: EMBL:X00970; NID:928325; PIDN:CA68784.1; PID:928326
 C:Genetics:
 A:Gene: GDB:ACR
 A:Cross-references: GDB:119645; OMIM:102480
 A:Map position: 22q13-22qter
 A:Introns: 26/2; 94/2; 189/2; 237/3
 C:Superfamily: acrosin; trypsin homology
 C:Keywords: glycoprotein; hydrolase; serine proteinase; sperm
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-421/Product: acrosin #status predicted <MAT>
 F:20-42/Product: acrosin light chain #status predicted <LCH>
 F:43-421/Product: acrosin heavy chain #status predicted <HCH>
 F:43-285/Domain: trypsin homology <TRY>
 F:302-379/Region: proline-rich
 F:22-210/Binding site: carbonylate (Aen) (covalent) #status predicted
 F:25-154/Disulfide bonds: #status predicted
 F:29-162/Disulfide bonds: #status predicted
 F:73-89/Disulfide bonds: #status predicted
 F:88-147, 240/Active site: His, Asp, Ser #status predicted
 F:177-246/Disulfide bonds: #status predicted
 F:209-225/Disulfide bonds: #status predicted
 F:236-266/Disulfide bonds: #status predicted
 Query Match 31.9%; Score 492; DB 1; Length 421;
 Best Local Similarity 40.5%; Pred. No. 8, 4e-37;
 Matches 106; Conservative 42; Mismatches 88; Indels 26; Gaps 10;

QY 254 SGCAKXRPVGVGNWVFTDMI 275
 DB 264 VGCALRRPQITATWPIYLMWI 285
 RESULT 13
 T30337
 polypeptide - African clawed frog
 C/Species: Xenopus laevis (African clawed frog)
 C/Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 03-Feb-2003
 C/Accession: T30337
 R:Yang, J.C.; Lindsay, L.L.; Hedrick, J.L.
 Submitted to the EMBL Data Library, March 1998
 A>Description: cDNA cloning of ovocytinase, a chymotrypsin-like protease released from
 A:Reference number: Z20829
 A:Accession: T30337
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1524 <YAN>
 A:Cross-references: EMBL:U81290; NID:92981640; PID:92981641; PIDN:AAC24717.1
 C:Superfamily: trypsin related polypeptide; trypsin homology
 Query Match 31.8%; Score 490; DB 2; Length 1524;
 Best Local Similarity 41.7%; Pred. No. 5, 9e-36;
 Matches 98; Conservative 43; Mismatches 84; Indels 10; Gaps 6;

QY 46 RIVGESALPGAMPQVSLHVNHCSSITPEWITVPAHCVEPLNPMWTAAGI 105
 DB 583 RIVGESALPGAMPQVSLHVNHCSSITPEWITVPAHCVEPLNPMWTAAGI 639
 QY 106 LKQSFMPYAGYQVEKVIS--HPNYSKTQNDIALMQLKQELTFNDLVKPCLPNPGM 162
 DB 640 -DNRMLNSTRQIRIKTIRLHDNYSSTYNDIALVLEPLDNDPVRPCLPEPSE 698
 QY 163 MLQPEQLCWISGMGAITEKGTSEVNAKVLITQRCSNRYVNDLTPAMICGFLQ 222
 DB 699 VLPSPVCAVTGNGTANBQPLGQLOQLPLDLSITNTST-YSGELTDHMLCAGFPS 757
 QY 223 G-NVDSQCGDSGGXLY-TSKNNIMWLIGDTSWCGCAKAYRPVGVGNWVFTDMI 275
 DB 758 SKEHACQDSGSLPVCQNEKQFSLYGLVSNBGGCRVSKGVYTKVLLFFTWI 812

RESULT 14
 A34170
 acrosin (BC 3.4.21.10) precursor - pig
 N:Alternate names: 53k fucose-binding protein
 C/Species: Sus scrofa domestica (domestic pig)
 C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C/Accession: A34170; S08994; S02428; S04940; S1657; S02780; S10695; S12968
 R:Baba, T.; Kashiwabara, S.; Watanabe, K.; Itoh, H.; Michikawa, Y.; Kimura, K.; Takada
 J. Biol. Chem. 264, 11920-11927, 1989
 A>Title: Activation and maturation mechanisms of boar acrosin zymogen based on the ded
 A:Reference number: A34170; MUID:89108595; PMID:2745422
 A:Accession: A34170
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-415 <BAB>
 A:Cross-references: GB:J04950; NID:9164702; PIDN:AAA1131.1; PID:9164703
 R:Kochova, D.; Toepfer-Petersen, E.; Zucker, A.; Jonakova, V.
 Biol. Chem. Hoppe-Seyler 371, 317-323, 1990
 A>Title: Is spermioxygen a modified proacrosin? Isolation, purification, and partial ch
 A:Reference number: S08994; MUID:90253655; PMID:2111146
 A:Accession: S08994
 A:Molecule type: protein
 A:Residues: 'X', 18, 'X', 20-25, 'X', 27-32, 'X', 34-38, 'X', 40-50 <CEC>
 R:Toepfer-Petersen, E.; Henschen, A.
 FEBS Lett. 226, 38-42, 1987
 A>Title: Acrosin shows zona and fucose binding, novel properties for a serine proteina
 A:Reference number: S02428; MUID:88083633; PMID:3480243
 A:Accession: S02428
 A:Molecule type: protein

PI Bandman O, Lal P;
XX
DR WPI; 2000-282523/24.
DR N-PSDB; AAA12975.
XX
PT Polynucleotide encoding human prostate-associated protease useful for
PT diagnosing and treating cancers, prostate disorders and
PT gastrointestinal disorders -
XX
PS Claim 1; Fig 1A-B; 27pp; English.
XX
XX This sequence represents human prostate-associated protease (HUPAP). cDNA
CC encoding HUPAP was initially identified in a spinal cord cDNA library.
CC the cDNA encoding this sequence representing a consensus of overlapping
CC and/or extended nucleic acid sequences from spinal cord, prostate tumour
CC and colon cDNA libraries. HUPAP is a serine protease with structural and
CC functional homology with bovine enterokinase, human pancreatic kallikrein
CC and African rat renal kallikrein, sharing 38% homology with bovine
CC enterokinase. In addition, HUPAP is related to prostate-specific antigen
CC (PSA), a kallikrein which is a highly sensitive marker for prostate
CC cancer. HUPAP and nucleic acids encoding it are useful for the diagnosis,
CC prevention and treatment of gastrointestinal disorders such as ulcerative
CC colitis, pancreatitis, cancers, and prostatic disorders. HUPAP agonists
CC may be used to treat gastrointestinal disorders, and HUPAP antagonists
CC and inhibitors may be used to suppress excessive cell proliferation,
CC which is of use in cancer therapy. HUPAP nucleic acids are also useful
CC for generating hybridisation probes which may be used for mapping.
CC naturally occurring genomic sequences.
XX
XX
SQ Sequence 283 AA;
XX
XX
Query Match 99.9%; Score 1541; DB 21; Length 283;
Best Local Similarity 100.0%; Pred. No. 3.7e-136;
Matches 283; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 1 MKLNTSAGNVDIYKKLYHSDACSSKAVYSLRCIAGVNLSSRSRIYVGSALPGAMPW 60
DB 1 MKLNTSAGNVDIYKKLYHSDACSSKAVYSLRCIAGVNLSSRSRIYVGSALPGAMPW 60
OY 61 QVSLHVNQNVHVCSSITTPENIVTAHCVEKPLNPPHMTAFAGILROSFMFYGAGYVE 120
DB 61 QVSLHVNQNVHVCSSITTPENIVTAHCVEKPLNPPHMTAFAGILROSFMFYGAGYVE 120
OY 121 KVISHPNYDSKTKKNDIALMKLQKPLTFNDLVKPVCLPNPQMTLOPEOLCNIISGKATGE 180
DB 121 KVISHPNYDSKTKKNDIALMKLQKPLTFNDLVKPVCLPNPQMTLOPEOLCNIISGKATGE 180
OY 181 KGKTSVTLNAKAVLLIFORCNSRYVDNLITPAMICAGFLQGVNDSQGSQGLVTSK 240
DB 181 KGKTSVTLNAKAVLLIFORCNSRYVDNLITPAMICAGFLQGVNDSQGSQGLVTSK 240
OY 241 NNIMWMLIGDTSWGSCKAKAYRPGYGVNVFTDVIYRQMRADG 283
DB 241 NNIMWMLIGDTSWGSCKAKAYRPGYGVNVFTDVIYRQMRADG 283
XX
RESULT 2
ABR78328
ID ABR78328 standard; Protein; 283 AA.
XX
XX ABR78328;
XX
DT 05-DEC-2002 (first entry)
XX
XX Amino acid sequence of human prostate-associated protease (HUPAP).
XX
XX Human; prostate-associated protease; HUPAP; prostatic disorder;
KM prostate cancer; benign prostatic hyperplasia; gastrointestinal disorder;
KM congenital enterokinase deficiency; pancreatitis; ulcerative colitis;
KM cancer; enzyme.
XX
XX Homo sapiens.
XX

PH Key Location/Qualifiers
FT Modified-site 4
FT FT /note= "potential N-glycosylation site"
FT Misc-difference 235
FT FT /note= "unspecified residue encoded by CNT"
XX
XX US2002119531-A1.
XX
XX 29-AUG-2002.
XX
XX 19-NOV-2001; 2001US-0988975.
XX
XX 27-FEB-1997; 97US-0807151.
XX 07-JAN-2000; 2000US-0478957.
XX
XX (BAND/) BANDMAN O.
XX (LALP/) LAL P G.
XX
XX Bandman O, Lal PG;
XX
XX WPI; 2002-627171/67.
XX N-PSDB; ABV72262.
XX
XX Novel human prostate-associated protease (HUPAP), useful for treating
PT gastrointestinal disorders such as congenital enterokinase deficiency
PT and for producing anti-HUPAP antibodies -
XX
XX
XX Claim 1; Fig 1A-B; 27pp; English.
XX
XX The present sequence represents human prostate-associated protease
CC (HUPAP). The antibody that specifically binds to HUPAP is useful in the
CC treatment of a prostatic (e.g. prostate cancer and benign prostatic
CC hyperplasia) or gastrointestinal disorder (e.g. congenital enterokinase
CC deficiency). The HUPAP antagonist is useful for treating the prostate
CC disorders described above. The HUPAP antagonist or antibody is also
CC useful for treating pancreatitis, ulcerative colitis and cancers of the
CC oesophagus, stomach, small intestine, large intestine or colon. HUPAP or
CC its agonist is useful for treating the gastrointestinal disorders
CC described above.
XX
XX
SQ Sequence 283 AA;
XX
XX
Query Match 99.9%; Score 1541; DB 23; Length 283;
Best Local Similarity 100.0%; Pred. No. 3.7e-136;
Matches 283; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 1 MKLNTSAGNVDIYKKLYHSDACSSKAVYSLRCIAGVNLSSRSRIYVGSALPGAMPW 60
DB 1 MKLNTSAGNVDIYKKLYHSDACSSKAVYSLRCIAGVNLSSRSRIYVGSALPGAMPW 60
OY 61 QVSLHVNQNVHVCSSITTPENIVTAHCVEKPLNPPHMTAFAGILROSFMFYGAGYVE 120
DB 61 QVSLHVNQNVHVCSSITTPENIVTAHCVEKPLNPPHMTAFAGILROSFMFYGAGYVE 120
OY 121 KVISHPNYDSKTKKNDIALMKLQKPLTFNDLVKPVCLPNPQMTLOPEOLCNIISGKATGE 180
DB 121 KVISHPNYDSKTKKNDIALMKLQKPLTFNDLVKPVCLPNPQMTLOPEOLCNIISGKATGE 180
OY 181 KGKTSVTLNAKAVLLIFORCNSRYVDNLITPAMICAGFLQGVNDSQGSQGLVTSK 240
DB 181 KGKTSVTLNAKAVLLIFORCNSRYVDNLITPAMICAGFLQGVNDSQGSQGLVTSK 240
OY 241 NNIMWMLIGDTSWGSCKAKAYRPGYGVNVFTDVIYRQMRADG 283
DB 241 NNIMWMLIGDTSWGSCKAKAYRPGYGVNVFTDVIYRQMRADG 283
XX
RESULT 3
AAE14615
ID AAE14615 standard; Protein; 283 AA.
XX
XX AAE14615;
XX

```
DT 18-JUN-2002 (first entry)
XX
XX Human prostate-associated protease (HUPAP).
DE Human prostate-associated protease; HUPAP; serine protease; kallikrein;
XX Human; prostate-associated protease; HUPAP; serine protease; kallikrein;
KW gastroduodenal disorder; pancreatic; ulcerative colitis; cancer;
KW adenocarcinoma; prostatic hyperplasia; prostate disorder.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
XX Region 22..45
XX /note= "Antigenic fragment. This region is
XX specifically claimed in claim 5 of the specification"
FT Active-site 87
FT /note= "Essential for serine protease activity"
FT Active-site 136
FT /note= "Essential for serine protease activity"
FT Active-site 232
FT /note= "Essential for serine protease activity"
FT Misc-difference 235
FT /label= Unknown
FT /note= "Encoded by CNT"
XX
XX US6350448-B1.
XX
XX 26-FEB-2002.
XX
XX 07-JAN-2000; 2000US-0478957.
XX
XX 27-FEB-1997; 97US-0807151.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Bandman O, Lal P;
XX
XX WPI; 2002-215263/27.
XX
XX N-PSDB; AAD27930.
XX
XX A purified human prostate-associated protease useful for treating a
XX gastroduodenal disorder, screening for a compound or molecule that
XX binds to it and for purifying such compounds -
XX
XX Claim 1; Fig 1; 27pp; English.
XX
XX The present sequence is human prostate-associated protease (HUPAP),
XX a serine protease belonging to kallikrein family. The HUPAP or its
XX agonist is used for treating gastroduodenal disorders (e.g.
XX pancreaticitis, ulcerative colitis). The HUPAP antagonist can be used to
XX treat cancer (e.g. adenocarcinoma, colon cancer), and prostate disorders
XX e.g. prostatic hyperplasia. The HUPAP polypeptide can also be
XX used to screen a library of compounds that bind to it.
XX
XX Sequence 283 AA;
SQ
Query Match 99.9%; Score 1541; DB 23; Length 283;
Best Local Similarity 100.0%; Pred. No. 3.7e-136;
Matches 283; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKLNTSAGNDVLYKLYSDACSSRAVVSIRCIACGVNLSRQSRIVGSGSLPGAPW 60
DB 1 MKLNTSAGNDVLYKLYSDACSSRAVVSIRCIACGVNLSRQSRIVGSGSLPGAPW 60
QY 61 QVSLHVQNVHVCSSGIITPEWITVAHCVKEPLNPMWHTAAGILRSGFMFYGAGYQYE 120
DB 61 QVSLHVQNVHVCSSGIITPEWITVAHCVKEPLNPMWHTAAGILRSGFMFYGAGYQYE 120
QY 121 KVISHPNVDSKTKNDILMLQOKPLTPDLYKPVCLPMPGMLQPEQICWISGKATSE 180
DB 121 KVISHPNVDSKTKNDILMLQOKPLTPDLYKPVCLPMPGMLQPEQICWISGKATSE 180
QY 181 KGKTSVANAARVLLIETQRCNSRYVDNLITPAMICAGFLQGNVDSGCGSGXIVTSK 240
DB 181 KGKTSVANAARVLLIETQRCNSRYVDNLITPAMICAGFLQGNVDSGCGSGXIVTSK 240
```

```
DB 181 KGKTSVANAARVLLIETQRCNSRYVDNLITPAMICAGFLQGNVDSGCGSGXIVTSK 240
QY 241 NNIMWLIQDTSWGGGCAKAYRPGVGNVWFETDMITRORADG 283
DB 241 NNIMWLIQDTSWGGGCAKAYRPGVGNVWFETDMITRORADG 283
RESULT 4
AAU69962
ID AAU69962 standard; Protein; 393 AA.
XX
XX AAU69962;
AC
XX
XX 30-JAN-2002 (first entry)
DT
XX
XX Human prostate cDNA encoded protein #88.
DE
XX
XX Human; prostate cancer; cytostatic; immunostimulant; tumour; immunogen.
XX
XX Homo sapiens.
OS
XX
XX WO200173032-A2.
XX
XX 04-OCT-2001.
XX
XX 27-MAR-2001; 2001WO-US09919.
XX
XX 27-MAR-2000; 2000US-0536857.
XX
XX 09-MAY-2000; 2000US-0568100.
XX
XX 12-MAY-2000; 2000US-0570737.
XX
XX 13-JUN-2000; 2000US-0593793.
XX
XX 27-JUN-2000; 2000US-0605783.
XX
XX 10-AUG-2000; 2000US-0636215.
XX
XX 29-AUG-2000; 2000US-0651236.
XX
XX 06-SEP-2000; 2000US-0657279.
XX
XX 02-OCT-2000; 2000US-0679426.
XX
XX 10-OCT-2000; 2000US-0685166.
XX
XX (CORI-) CORIXA CORP.
XX
XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kaloe MD;
XX Fanger GR, Reiter MW, Stolk JA, Day CH, Veddyck TS, Carter D;
XX Li SX, Wang A, Skelky YAM, Hepler WT, Henderson RA;
XX
XX WPI; 2001-639232/73.
XX
XX New human prostate-specific polypeptides and polynucleotides useful for
XX the diagnosis and treatment of cancer, especially prostate cancer -
XX
XX Claim 2; Page 574-575; 579pp; English.
XX
XX The invention relates to isolated prostate-specific
XX polynucleotides, polypeptides, fusion proteins of the polypeptides,
XX antibodies raised against the polypeptides (or antigenic epitopes
XX derived from them) and antigen-presenting cells expressing the
XX polypeptides. The antibodies are useful for detecting the presence of
XX cancer, especially prostate cancer. The polypeptides, polynucleotides and
XX the antigen-presenting cells are useful for stimulating and/or expanding
XX T cells specific for a tumour protein, and for inhibiting the development
XX of cancer especially prostate cancer. Compositions comprising the
XX polynucleotide and/or polypeptide are useful for stimulating an immune
XX response, and for treating cancer. The oligonucleotide is useful for
XX detecting cancer. The present sequence is a prostate specific
XX polypeptide of the invention.
XX
XX Sequence 393 AA;
SQ
Query Match 99.8%; Score 1540; DB 22; Length 393;
Best Local Similarity 99.6%; Pred. No. 7e-116;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MKLNTSAGNDVLYKLYSDACSSRAVVSIRCIACGVNLSRQSRIVGSGSLPGAPW 60
DB 1 MKLNTSAGNDVLYKLYSDACSSRAVVSIRCIACGVNLSRQSRIVGSGSLPGAPW 60
```

Db 111 MKLNTSAGNVDIYKLLYHS DACSSKAVYSLRCIACGVNLNSRSQSRIVGGSALPGAMP 170
Qy 61 QVSLHVQNVHVC GSGIITPEWIVTAAHCYKELNPNWMTAFAGILROSFMFYAGAYQE 120
Db 171 QVSLHVQNVHVC GSGIITPEWIVTAAHCYKELNPNWMTAFAGILROSFMFYAGAYQE 230
Qy 121 KVISHPNVD SKTKNNDIAMKLOKPLTFNDLVKPVCLPMPGMLLOPEOLCMTSGMGATEE 180
Db 231 KVISHPNVD SKTKNNDIAMKLOKPLTFNDLVKPVCLPMPGMLLOPEOLCMTSGMGATEE 290
Qy 181 KGTSEVINA AAVLLIETORCNSRYVDNLITPAMI CAGFLQGNVDS CGDSGGLVTSK 240
Db 291 KGTSEVINA AAVLLIETORCNSRYVDNLITPAMI CAGFLQGNVDS CGDSGGLVTSK 350
Qy 241 NNIMWLIGDTSW GSGCAKAYRPGVYGNVVFPTDWIYRQRADG 283
Db 351 NNIMWLIGDTSW GSGCAKAYRPGVYGNVVFPTDWIYRQRADG 393

RESULT 5
AAM01317
ID AAM01317 standard; Protein; 393 AA.

AC AAM01317;

DT 04-OCT-2001 (first entry)

DE P1000C partial amino acid sequence 100-492.

KM Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW cytosstatic; gene therapy; metastasis.

XX Homo sapiens.

XX PN. WO200151633-AZ.

XX PD 19-JUL-2001.

XX PP 16-JAN-2001; 2001WO-US01574.

XX PR 14-JAN-2000; 2000US-0483672.

XX PA (CORI-) CORIXA CORP.

XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG,
PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAM,
PI Wang A, Meagher MJ;
XX WPI, 2001-425673/45.

XX New polynucleotide encoding a prostate-specific protein, for
PT diagnosing, monitoring and treating prostate cancer in a patient and
PT for use in vaccines -
XX

PS Claim 2; Page 540-541; 543pp; English.

XX The present invention describes polynucleotide sequences (I) which encode
CC prostate-specific proteins (II). (I) and (II) have cytostatic activity.
CC and can be used in vaccine production and gene therapy. (I), (II),
CC antibodies to (II), fusion proteins comprising (II), and isolated
CC T cells prepared using (I) or (II) are used treat cancer in a patient.
CC (I) and the antibodies are also used in the detection of cancer in a
CC patient. The cancer that is diagnosed or treated is particularly
CC prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
CC (I) can be used for monitoring the progression of cancer in a patient.
CC (I) and (II) can also be used to improve diagnostic and therapeutic
CC methods for prostate cancer. They can indicate the level of metastasis
CC as well as the prostate volume. AAH93357 to AAH93944 and AAM01115 to
CC AAM01318 represent polynucleotide and amino acid sequences used in the
CC exemplification of the present invention.

XX Sequence 393 AA;
SQ

Query Match 99.8%; Score 1540; DB 22; Length 393;
Best Local Similarity 99.6%; Pred. No. 7e-1156; 1; Indels 0; Gaps 0;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MKLNTSAGNVDIYKLLYHS DACSSKAVYSLRCIACGVNLNSRSQSRIVGGSALPGAMP 60
Db 111 MKLNTSAGNVDIYKLLYHS DACSSKAVYSLRCIACGVNLNSRSQSRIVGGSALPGAMP 170
Qy 61 QVSLHVQNVHVC GSGIITPEWIVTAAHCYKELNPNWMTAFAGILROSFMFYAGAYQE 120
Db 171 QVSLHVQNVHVC GSGIITPEWIVTAAHCYKELNPNWMTAFAGILROSFMFYAGAYQE 230
Qy 121 KVISHPNVD SKTKNNDIAMKLOKPLTFNDLVKPVCLPMPGMLLOPEOLCMTSGMGATEE 180
Db 231 KVISHPNVD SKTKNNDIAMKLOKPLTFNDLVKPVCLPMPGMLLOPEOLCMTSGMGATEE 290
Qy 181 KGTSEVINA AAVLLIETORCNSRYVDNLITPAMI CAGFLQGNVDS CGDSGGLVTSK 240
Db 291 KGTSEVINA AAVLLIETORCNSRYVDNLITPAMI CAGFLQGNVDS CGDSGGLVTSK 350
Qy 241 NNIMWLIGDTSW GSGCAKAYRPGVYGNVVFPTDWIYRQRADG 283
Db 351 NNIMWLIGDTSW GSGCAKAYRPGVYGNVVFPTDWIYRQRADG 393

RESULT 6

ABB95422
ID ABB95422 standard; Protein; 393 AA.

AC ABB95422;

DT 19-JUL-2002 (first entry)

DE Human P1000C fragment SEQ ID NO 934.

KM Human; cancer; prostate cancer; vaccine; cytosstatic; immunostimulant;
KW gene therapy.

XX Homo sapiens.

XX PN US200202248-A1.

XX PD 21-FEB-2002.

XX PP 12-JAN-2001; 2001US-0759143.

XX PR 25-FEB-1997; 97US-0806099.

XX PR 01-AUG-1997; 97US-0904804.

XX PR 09-FEB-1998; 98US-0020956.

XX PR 25-FEB-1998; 98US-0030607.

XX PR 14-JUL-1998; 98US-0115453.

XX PR 23-SEP-1998; 98US-0159812.

XX PR 15-JAN-1999; 99US-0232149.

XX PR 09-APR-1999; 99US-0288946.

XX PR 13-JUL-1999; 99US-0352616.

XX PR 12-NOV-1999; 99US-0439313.

XX PR 18-NOV-1999; 99US-0443686.

XX PR 14-JAN-2000; 2000US-0483672.

XX PR 27-MAR-2000; 2000US-0536857.

XX PR 09-MAY-2000; 2000US-0568107.

XX PR 12-MAY-2000; 2000US-0570737.

XX PR 13-JUN-2000; 2000US-0593793.

XX PR 27-JUN-2000; 2000US-0605783.

XX PR 10-AUG-2000; 2000US-0636215.

XX PR 29-AUG-2000; 2000US-0651236.

XX PR 06-SEP-2000; 2000US-0657279.

XX PR 02-OCT-2000; 2000US-0679426.

XX PR 10-OCT-2000; 2000US-0685166.

XX (XUJ/) XU J.
XX (DILL/) DILLON D C.
XX (MITC/) MITCHAM J L.
XX (HARL/) HARLOCKER S L.

PA (JIAN/) JIANG Y.
 PA (KALO/) KALOS M D.
 PA (FANG/) FANGER G R.
 PA (RETT/) RETTER M W.
 PA (STOL/) STOLK J A.
 PA (DAYC/) DAY C H.
 PA (VEDV/) VEDVICK T S.
 PA (CART/) CARTER D.
 PA (LISX/) LI S X.
 PA (WANG/) WANG A.
 PA (SKEI/) SKEIKY Y A W.
 PA (HEPL/) HEPLER W T.
 PA (HEND/) HENDERSON R A.
 XX
 XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
 PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
 PI Li SX, Wang A, Skeiky YAM, Hepler WT, Henderson RA;
 XX WPI; 2002-255649/30.
 DR
 XX New prostate-specific polynucleotides for diagnosing and treating
 PT diseases, in particular prostate cancer, and as markers for the
 PT progression of cancer -
 XX
 PS Claim 2; SEQ ID NO 934; 87bp; English.
 XX
 CC The present invention provides prostate-specific coding sequences and
 CC their encoded proteins. These can be used in the diagnosis and treatment
 CC of cancers, particularly prostate cancer. The present sequence is a
 CC protein described in the invention.
 XX
 SQ Sequence 393 AA;
 XX
 Query Match 99.8%; Score 1540; DB 23; Length 393;
 Best Local Similarity 99.6%; Pred. No. 7e-136;
 Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 MKLNTSAGNVDIYKLYHSDACSSKAVVSLRCLACGVNLNSSRQSRIVGSESLPGAMPW 60
 Db 111 MKLNTSAGNVDIYKLYHSDACSSKAVVSLRCLACGVNLNSSRQSRIVGSESLPGAMPW 170
 Oy 61 OVSILHVQNVHVCQGSITTPMTVTAACHCEKPLNPMWMTAFAGILROSFMFYAGAYOVE 120
 Db 171 QVSLHVQNVHVCQGSITTPMTVTAACHCEKPLNPMWMTAFAGILROSFMFYAGAYOVE 230
 Oy 121 KYISHPNVYSKTKNDIALMKLOKPTFNDLVYKVCLEPNFGMLQPEQLCWMISGMATTEE 180
 Db 231 KYISHPNVYSKTKNDIALMKLOKPTFNDLVYKVCLEPNFGMLQPEQLCWMISGMATTEE 290
 Oy 181 KQKTSFVLNAKAVLLETORCNSRYVYDNLITPAMT CAGFLQGNVNSCCGDSGGLVYTSK 240
 Db 291 KQKTSFVLNAKAVLLETORCNSRYVYDNLITPAMT CAGFLQGNVNSCCGDSGGLVYTSK 350
 Oy 241 NNIMWLIGDTSWGSQCAKAYRPGVYGNVAVFTDVIYRORADG 283
 Db 351 NNIMWLIGDTSWGSQCAKAYRPGVYGNVAVFTDVIYRORADG 393
 RESULT 7
 ABU71853
 ID ABU71853 standard; Protein; 393 AA.
 XX
 AC ABU71853;
 XX
 DT 10-JUN-2003 (first entry)
 XX
 XX Prostate cancer associated protein #69.
 DB Prostate cancer; vaccine; gene therapy; cytostatic; fusion protein;
 KW immunogen; cancer; prostate specific antigen; PSA;
 KW prostatic acid phosphatase; PAP; prostate specific membrane antigen;
 KM PSMA.
 XX

OS Homo sapiens.
 XX
 XX us2002192763-A1.
 PN
 XX
 PD 19-DEC-2002.
 XX
 XX 29-JUN-2001; 2001US-0895793.
 PF
 XX 17-APR-2000; 2000US-157455P.
 PR 04-OCT-2000; 2000US-0679272.
 PR 28-MAR-2001; 2001US-0822827.
 XX
 XX (XUJ/) XU J.
 PA (DILL/) DILLON D C.
 PA (MITC/) MITCHAM J L.
 PA (HARL/) HARLOCKER S L.
 PA (JIAN/) JIANG Y.
 PA (KALO/) KALOS M D.
 PA (FANG/) FANGER G R.
 PA (RETT/) RETTER M W.
 PA (STOL/) STOLK J A.
 PA (DAYC/) DAY C H.
 PA (VEDV/) VEDVICK T S.
 PA (CART/) CARTER D.
 PA (LISX/) LI S X.
 PA (WANG/) WANG A.
 PA (SKEI/) SKEIKY Y A W.
 PA (HEPL/) HEPLER W T.
 PA (HEND/) HENDERSON R A.
 PA (HURA/) HURAL J.
 PA (MCNE/) MCNEILL P D.
 PA (HOUN/) HOUGHTON R L.
 PA (DEAS/) Y DE BASSELS C V.
 PA (FOYT/) FOY T M.
 XX
 XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
 PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
 PI Li SX, Wang A, Skeiky YAM, Hepler WT, Henderson RA, Hurai J;
 PI McNeill PD, Houghton RL, Y De Bassols CV, Foy TM;
 XX WPI; 2003-352711/33.
 DR
 XX
 XX New fusion protein comprising prostate-specific polypeptides, or its
 PT immunogenic portions, useful for diagnosing, preventing and/or treating
 PT cancer, particularly prostate cancer -
 XX
 PS Example 2; SEQ ID NO 934; 85bp; English.
 PS
 XX
 CC The invention describes a fusion protein comprising at least one amino
 CC acid sequence of immunogenic portions of any of the 3 sequences not
 CC defined in the specification, or sequences having at least 70 or 90 %
 CC sequence identity to any one of the 35 sequences defined in the USPTO
 CC web site, which is encoded by any of the 4 nucleotide sequences not
 CC defined in the specification. The fusion protein, composition and
 CC methods are useful for diagnosing, preventing and/or treating cancer,
 CC particularly prostate cancer. The proteins are useful as markers to
 CC indicate the presence or absence of cancer. This is the amino acid
 CC sequence of a prostate cancer therapy associated protein.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from the US patent office at
 CC seqdata.uspto.gov/sequence.html?DocID=US20020192763.
 CC
 XX
 SQ Sequence 393 AA;
 XX
 Query Match 99.8%; Score 1540; DB 24; Length 393;
 Best Local Similarity 99.6%; Pred. No. 7e-136;
 Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 MKLNTSAGNVDIYKLYHSDACSSKAVVSLRCLACGVNLNSSRQSRIVGSESLPGAMPW 60
 Db 111 MKLNTSAGNVDIYKLYHSDACSSKAVVSLRCLACGVNLNSSRQSRIVGSESLPGAMPW 170

QY 61 QVSLHVNQVHVGGSIIITPEWIVTAHCVKEPLNPMHTAFAGILROSFMFYAGYQVE 120
DB 171 QVSLHVNQVHVGGSIIITPEWIVTAHCVKEPLNPMHTAFAGILROSFMFYAGYQVE 230
QY 121 KVISHPNVDSTKKNNDIALMKLOKPLTENDLVKPVCLPNPGMMLQPEOLCMTISGMGATEE 180
DB 221 KVISHPNVDSTKKNNDIALMKLOKPLTENDLVKPVCLPNPGMMLQPEOLCMTISGMGATEE 290
QY 181 KGTSEVTAANAAYVLLIETORCNSRYVDNLITPAMI CAGFLQGNVDS CGDSGGGLVTSK 240
DB 291 KGTSEVTAANAAYVLLIETORCNSRYVDNLITPAMI CAGFLQGNVDS CGDSGGGLVTSK 350
QY 241 NNIMWLIQDTSWGS GCAKAYRPGVGNVWFTDMITRQMRADG 283
DB 351 NNIMWLIQDTSWGS GCAKAYRPGVGNVWFTDMITRQMRADG 393

RESULT 8
AA92050
ID AA92050 standard; Protein; 492 AA.
AC AAY92050;
XX
DT 01-AUG-2000 (first entry)
XX
DE HRPcA6/7 polypeptide from androgen-inducible gene clone.
XX
KW Androgen inducible; testosterone; prostate cancer; cytostatic;
KW TMPRSS2; diagnosis.
XX
OS Homo sapiens.
XX
FN WO200018961-A2.
XX
PD 06-APR-2000.
XX
PE 30-SEP-1999; 99WO-US22535.
XX
PR 30-SEP-1998; 98US-0163759.
XX
PR 30-SEP-1998; 98US-0164159.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Macbeth KJ. Shyjan AM;
XX
DR WPI; 2000-293182/25.
XX
DR N-PSDB; AAA08803.
XX
PT Novel methods for identifying compounds for treating prostate cancer
PT comprising measuring the level of expression or activity of 1 or more
PT of 11 genes or their products
XX
PS Claim 2; Fig 3; 108pp; English.
XX
CC This protein is encoded by a gene which is androgen (e.g. testosterone)
CC inducible in androgen-dependent prostate cancer cells (e.g. LNCaP cells)
CC and constitutively expressed in androgen-independent prostate cancer
CC cells (e.g. LNCaP cells). Agents which decrease the expression or
CC activity of these clones may slow or arrest the growth of prostate cancer
CC cells or may kill them. HRPcA6/7 can be obtained from the sequence of
CC the known gene for TMPRSS2. A compound useful for treating prostate
CC cancer can be identified in a novel method comprising measuring the
CC expression level, or activity, of HRPcA2, 3, 6/7, 8, 9, 10, 13, 14, 15,
CC 19, or peripheral-type benzodiazepine receptor (PBR) in a cell, in the
CC presence and absence of a test compound. The sequences may also be used
CC in diagnosis of prostate cancer and to determine efficacy of treatment
CC for prostate cancer.
XX
SQ Sequence 492 AA;

Query Match 99.8%; Score 1540; DB 21; Length 492;
Best Local Similarity 99.6%; Pred. No. 9.2e-136;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGLNLSAGVNDIYKRLYHSDACSKAVVSLRCLACGVNLNSRSRQRIYGSALPGAWPW 60
DB 210 MGLNLSAGVNDIYKRLYHSDACSKAVVSLRCLACGVNLNSRSRQRIYGSALPGAWPW 269
QY 61 QVSLHVNQVHVGGSIIITPEWIVTAHCVKEPLNPMHTAFAGILROSFMFYAGYQVE 120
DB 270 QVSLHVNQVHVGGSIIITPEWIVTAHCVKEPLNPMHTAFAGILROSFMFYAGYQVE 329
QY 121 KVISHPNVDSTKKNNDIALMKLOKPLTENDLVKPVCLPNPGMMLQPEOLCMTISGMGATEE 180
DB 330 KVISHPNVDSTKKNNDIALMKLOKPLTENDLVKPVCLPNPGMMLQPEOLCMTISGMGATEE 389
QY 181 KGTSEVTAANAAYVLLIETORCNSRYVDNLITPAMI CAGFLQGNVDS CGDSGGGLVTSK 240
DB 390 KGTSEVTAANAAYVLLIETORCNSRYVDNLITPAMI CAGFLQGNVDS CGDSGGGLVTSK 449
QY 241 NNIMWLIQDTSWGS GCAKAYRPGVGNVWFTDMITRQMRADG 283
DB 450 NNIMWLIQDTSWGS GCAKAYRPGVGNVWFTDMITRQMRADG 492

RESULT 9
AAY7726
ID AAY7726 standard; Protein; 492 AA.
AC AAY7726;
XX
DT 12-MAY-2000 (first entry)
XX
DE Human tumour suppressor TMPRSS2 polypeptide.
XX
KW Tumour suppressor gene; TMPRSS2; cancer; human; drug design;
KW gene therapy; protein therapy.
XX
OS Homo sapiens.
XX
FN WO20000605-A1.
XX
PD 06-JAN-2000.
XX
PE 29-JUN-1999; 99WO-US14622.
XX
PR 29-JUN-1998; 98US-0091044.
XX
PA (MYRI-) MYRIAD GENETICS INC.
XX
PI Wong AKC, Tavtigian SV, Teng DHF;
XX
DR WPI; 2000-170914/15.
XX
DR N-PSDB; AA287786.
XX
PT Novel tumor suppressor TMPRSS2 used for the diagnosis and prognosis of
PT human cancer -
XX
PS Claim 55; Page 77-79; 89pp; English.
XX
CC The invention provides a new tumour suppressor gene, designated TMPRSS2.
CC The TMPRSS2 polynucleotides and polypeptides can be used in methods for
CC diagnosing and prognosing predisposition to cancer in humans. The
CC polypeptides may also be used in assays to screen for compounds with
CC anti-cancer or therapeutic properties. The polypeptides are also useful
CC for rational drug design. The TMPRSS2 polynucleotides and polypeptides
CC may be used for gene therapy and protein therapy. The present sequence
CC represents the TMPRSS2 polypeptide.
XX
SQ Sequence 492 AA;

Query Match 99.8%; Score 1540; DB 21; Length 492;
Best Local Similarity 99.6%; Pred. No. 9.2e-136;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGLNLSAGVNDIYKRLYHSDACSKAVVSLRCLACGVNLNSRSRQRIYGSALPGAWPW 60

Db 210 MCLNTSAGVNDLYKCLYHSDACSSRAVSLRCLACGVNINSSRSRIVGESALPGAMPW 269

Qy 61 QVSLHVQNVHVCSSGSLITPEMTVTAHCVCEKPLANNPMHTAFAGILRQSFMYGAGYQVE 120

Db 270 QVSLHVQNVHVCSSGSLITPEMTVTAHCVCEKPLANNPMHTAFAGILRQSFMYGAGYQVE 329

Qy 121 KVISHPNVDSTKKNNDIALMKTOKPLTFNDLVKPVCLPMPGMLQPEOLCMISGWGATEE 180

Db 330 KVISHPNVDSTKKNNDIALMKTOKPLTFNDLVKPVCLPMPGMLQPEOLCMISGWGATEE 389

Qy 181 KGKTSBVNAAKAVLLIETQRCNSRYVDNLITPAMICAGFLQGNVDSGGSGGKLVTSK 240

Db 390 KGKTSBVNAAKAVLLIETQRCNSRYVDNLITPAMICAGFLQGNVDSGGSGGKLVTSK 449

Qy 241 NNIMWLIGDTSWGGCAKAYRPGYGVNVMFTDWTYRQRADG 283

Db 450 NNIMWLIGDTSWGGCAKAYRPGYGVNVMFTDWTYRQRADG 492

RESULT 10

AA44406 standard; Protein; 492 AA.

AA44406;

22-MAR-2000 (first entry)

Human 20P1F12-GTCT protein.

20P1F12; TMPRSS2; androgen; serine protease; 20P1F12-GTCT; cancer; transmembrane protein; colon; prostate; prostate tumour.

Homo sapiens.

MO9962942-A2.

09-DEC-1999.

01-JUN-1999; 99WO-US12253.

01-JUN-1998; 98US-0087598.

29-JUN-1998; 98US-0091474.

14-APR-1999; 99US-0129521.

(UBOCS-) UROGENESIS INC.

(AFAR/) AFAR D E.

(HUBB/) HUBERT R S.

(LEON/) LEONG K.

(RAIT/) RAITANO A B.

(SAFF/) SAFFRAN D C.

Afar DE, Hubert RS, Leong K, Raitano AB, Saffran DC;

WPI; 2000-116363/10.

N-PSDB; AA229636.

Novel cell surface antigen useful to treat colon and prostate cancer -

Claim 1; Fig 1; 58pp; English.

The present sequence is the 20P1F12 protein (also known as the TMPRSS2 protein) which is a prostate-specific, androgen-regulated, cell surface serine protease. It is a glycosylated type II transmembrane protein with an extracellular C-terminal serine protease domain, a scavenger receptor cysteine-rich domain, an LDL receptor class A domain and a predicted transmembrane domain. Host cells can be transformed to produce this protein, using vector containing 20P1F12/TMPRSS2 gene (also designated 20P1F12-GTCT), as deposited with ATCC accession number 2070977.

Anti-20P1F12/TMPRSS2 antibodies may be used as therapeutic agent for prostate and colon cancers, to image prostate cancer cells and prostate tumours, to identify ligands and cellular constituents that bind to a 20P1F12/TMPRSS2 gene product and for use as cancer vaccines.

XX SQ Sequence 492 AA;

Query Match 99.8%; Score 1540; DB 21; Length 492;

Best Local Similarity 99.6%; Pred. No. 9, 2e-136;

Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MCLNTSAGVNDLYKCLYHSDACSSRAVSLRCLACGVNINSSRSRIVGESALPGAMPW 60

Db 210 MCLNTSAGVNDLYKCLYHSDACSSRAVSLRCLACGVNINSSRSRIVGESALPGAMPW 269

Qy 61 QVSLHVQNVHVCSSGSLITPEMTVTAHCVCEKPLANNPMHTAFAGILRQSFMYGAGYQVE 120

Db 270 QVSLHVQNVHVCSSGSLITPEMTVTAHCVCEKPLANNPMHTAFAGILRQSFMYGAGYQVE 329

Qy 121 KVISHPNVDSTKKNNDIALMKTOKPLTFNDLVKPVCLPMPGMLQPEOLCMISGWGATEE 180

Db 330 KVISHPNVDSTKKNNDIALMKTOKPLTFNDLVKPVCLPMPGMLQPEOLCMISGWGATEE 389

Qy 181 KGKTSBVNAAKAVLLIETQRCNSRYVDNLITPAMICAGFLQGNVDSGGSGGKLVTSK 240

Db 390 KGKTSBVNAAKAVLLIETQRCNSRYVDNLITPAMICAGFLQGNVDSGGSGGKLVTSK 449

Qy 241 NNIMWLIGDTSWGGCAKAYRPGYGVNVMFTDWTYRQRADG 283

Db 450 NNIMWLIGDTSWGGCAKAYRPGYGVNVMFTDWTYRQRADG 492

RESULT 11

AAU69960 standard; Protein; 492 AA.

AAU69960;

30-JAN-2002 (first entry)

Human prostate cDNA encoded protein #86.

Human; prostate cancer; cytostatic; immunostimulant; tumour; immunogen.

Homo sapiens.

WO200173032-A2.

04-OCT-2001.

27-MAR-2001; 2001WO-US09919.

27-MAR-2000; 2000US-0536857.

09-MAY-2000; 2000US-0568100.

12-MAY-2000; 2000US-0570737.

13-JUN-2000; 2000US-0593793.

27-JUN-2000; 2000US-0605783.

10-AUG-2000; 2000US-0636215.

29-AUG-2000; 2000US-0651236.

06-SEP-2000; 2000US-0651279.

02-OCT-2000; 2000US-0679426.

10-OCT-2000; 2000US-0685166.

(CORI-) CORIXA CORP.

Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD, Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D, Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;

WPI; 2001-639232/73.

N-PSDB; AAS64178; AAS64179; AAS64180.

New human prostate-specific polypeptides and polynucleotides useful for the diagnosis and treatment of cancer, especially prostate cancer -

Claim 2; Page 573-574; 579pp; English.

PR 23-SEP-1998; 98US-0159812.
PR 15-JAN-1999; 99US-0232149.
PR 09-APR-1999; 99US-0288946.
PR 13-JUL-1999; 99US-0352616.
PR 12-NOV-1999; 99US-0439313.
PR 18-NOV-1999; 99US-0443686.
PR 14-JAN-2000; 2000US-0483672.
PR 27-MAR-2000; 2000US-0536857.
PR 09-MAY-2000; 2000US-0568100.
PR 13-JUN-2000; 2000US-0570737.
PR 27-JUN-2000; 2000US-0605783.
PR 10-AUG-2000; 2000US-0636215.
PR 29-AUG-2000; 2000US-0651236.
PR 06-SEP-2000; 2000US-0657279.
PR 02-OCT-2000; 2000US-0679426.
PR 10-OCT-2000; 2000US-0685166.
XX
XX (XUJ/) XU J.
PA (DILL/) DILLON D C.
PA (MITC/) MITCHAM J L.
PA (HARL/) HARLOCKER S L.
PA (JIANG/) JIANG Y.
PA (KALO/) KALOS M D.
PA (FANG/) FANGER G R.
PA (RETT/) RETTER M W.
PA (STOL/) STOLK J A.
PA (DAYC/) DAY C H.
PA (VEDV/) VEDVICK T S.
PA (CART/) CARTER D.
PA (LISK/) LI S X.
PA (WANG/) WANG A.
PA (SKEI/) SKEIKY Y A W.
PA (HEPL/) HEPLER W T.
PA (HEND/) HENDERSON R A.
XX
XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI Li SX, Wang A, Skeiky YAM, Hepler WT, Henderson RA;
XX
XX MPI; 2002-255649/30.
XX
XX New prostate-specific polynucleotides for diagnosing and treating
PT diseases, in particular prostate cancer, and as markers for the
PT progression of cancer
XX
XX Claim 2; SEQ ID NO 932; 87bp; English.
XX
XX The present invention provides prostate-specific coding sequences and
CC their encoded proteins. These can be used in the diagnosis and treatment
CC of cancers, particularly prostate cancer. The present sequence is a
CC protein described in the invention.
XX
XX
SQ Sequence 492 AA;
Query Match 99.8%; Score 1540; DB 23; Length 492;
Best Local Similarity 99.6%; Pred. No. 9.2e-136;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MCLNTSAGNVDTYKLYHSDACSSKAVSLRCLACGVNINSSRQSHIVGESALPGAMPW 60
DB 210 MCLNTSAGNVDTYKLYHSDACSSKAVSLRCLACGVNINSSRQSHIVGESALPGAMPW 269
QY 61 QVSLHVNQNVHVGCGSITTEPMVITAAHCYCEKPLNNFMHMTAFAGILRQSFMYGAGYOYE 120
DB 270 QVSLHVNQNVHVGCGSITTEPMVITAAHCYCEKPLNNFMHMTAFAGILRQSFMYGAGYOYE 329
QY 121 KVISHPNYSKTKNDIALMKLOKPLTFNDIVPVCLPBGWMLQREQLCWISGKATRE 180
DB 330 KVISHPNYSKTKNDIALMKLOKPLTFNDIVPVCLPBGWMLQREQLCWISGKATRE 389
QY 181 KKTSEVINAARVLLIFTORCNSRYVDNLITPAMICAGFLQGNVDSOQDSGGLVTSK 240

DB 390 KKTSEVINAARVLLIFTORCNSRYVDNLITPAMICAGFLQGNVDSOQDSGGLVTSK 449
QY 241 NNIMWLIGDTSWGSCKAAYRPGYGNMVFETDWTYRQMRADG 283
DB 450 NNIMWLIGDTSWGSCKAAYRPGYGNMVFETDWTYRQMRADG 492
RESULT 14
AAE18096
ID AAE18096 standard; Protein; 492 AA.
XX
XX AAE18096;
AC
XX 07-MAY-2002 (first entry)
DT
XX Human 20P1F12-GTCL protein.
DE
XX
XX Serine protease; 20P1F12/TMPRSS2; 20P1F12-GTCL; cell growth; neoplasm;
KW cancer; vaccine; human.
XX
XX Homo sapiens.
OS
XX MO200204953-A2.
XX
XX 17-JAN-2002.
XX
XX 12-JUL-2001; 2001WO-US22168.
XX
XX 12-JUL-2000; 2000US-0615285.
XX
XX (AGEN-) AGENSYS INC.
XX
XX Saferran D, Raitano AB, Hubert RS, Jakobovits A, Paris M;
PI Chailita-Bid FM;
PI
XX MPI; 2002-154967/20.
DR N-PSDB; AAD28778.
XX
XX Examining a biological sample for evidence of dysregulated cellular
PT growth, comprises comparing the status of prostate-specific,
PT androgen-regulated, secreted serine protease, 20P1F12/TMPRSS2, in a
PT corresponding normal sample
XX
XX
PS Claim 40; Fig 1; 161bp; English.
XX
XX The present invention relates to methods and compositions for the
CC diagnosis and therapy of prostate, colon, bladder, lung, ovarian and
CC kidney cancer derived from or based on a normally prostate-specific,
CC androgen regulated, cell membrane associated secreted serine protease
CC termed 20P1F12/TMPRSS2. The invention further relates to a method of
CC examining a biological sample for evidence of dysregulated cellular
CC growth comprises comparing the status of 20P1F12/TMPRSS2 gene (also
CC designated 20P1F12-GTCL) in the sample to the status of 20P1F12/TMPRSS2
CC in a corresponding normal sample. The invention also relates to 20P1F12/
CC TMPRSS2 polynucleotides and their corresponding proteins. Methods of the
CC invention are used for examining a sample such as blood, serum, stool,
CC urine, semen, or biopsy tissue for evidence of dysregulated cell growth.
CC The dysregulated cell growth is indicative of bladder cancer, lung
CC cancer, kidney cancer or ovarian cancer. It is useful for identifying
CC evidence of a neoplasm in a sample. Vaccines comprising an immunogenic
CC portion of 20P1F12/TMPRSS2 are useful for inhibiting growth of a cell
CC expressing 20P1F12/TMPRSS2 in a patient suffering from bladder cancer,
CC lung cancer, ovarian cancer or metastatic cancer. The present sequence
CC is human 20P1F12-GTCL protein.
XX
XX
SQ Sequence 492 AA;
Query Match 99.8%; Score 1540; DB 23; Length 492;
Best Local Similarity 99.6%; Pred. No. 9.2e-136;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MCLNTSAGNVDTYKLYHSDACSSKAVSLRCLACGVNINSSRQSHIVGESALPGAMPW 60

Db 210 MKLNTSAGNVDIYKLYHSDACSSKAVSLRCLACGVNINSSRSRIVGSESALPGAMPW 269
Qy 61 QVSLHVQNVHVCSSGSIITPEMIVTAHCEKPLNPMHTAFAGILROSFMFYGAGYQVE 120
Db 270 QVSLHVQNVHVCSSGSIITPEMIVTAHCEKPLNPMHTAFAGILROSFMFYGAGYQVE 329
Qy 121 KVISHPNVDSTKKNNDIALMKLOKPLTFNDLVKPVCLPNEGMLQPEQLCMISGMGATEE 180
Db 330 KVISHPNVDSTKKNNDIALMKLOKPLTFNDLVKPVCLPNEGMLQPEQLCMISGMGATEE 389
Qy 181 KGMTSEVLNAAKVLLIETORCNSRYVDNLITPAMI CAGFLQGNVDS CGDSCGXLVTSK 240
Db 330 KGMTSEVLNAAKVLLIETORCNSRYVDNLITPAMI CAGFLQGNVDS CGDSCGXLVTSK 449
Qy 241 NNIMWLIGDTSWGSCKAKAYRPGVGNVVFDTMIYRORADG 283
Db 450 NNIMWLIGDTSWGSCKAKAYRPGVGNVVFDTMIYRORADG 492

RESULT 15
ABU71851
ID ABU71851 standard; Protein; 492 AA.
AC ABU71851;
XX
DT 10-JUN-2003 (first entry)
XX
DE Prostate cancer associated protein #67.
XX
KM Prostate cancer; vaccine; gene therapy; cytostatic; fusion protein;
KM immunogen; cancer; prostate specific antigen. PSA;
KM prostatic acid phosphatase; PAP; prostate specific membrane antigen;
KM PSMa.
XX
OS Homo sapiens.
XX
PN US2002192763-A1.
XX
PD 19-DEC-2002.
XX
PF 29-JUN-2001; 2001US-0895793.
XX
PR 17-APR-2000; 2000US-157455P.
PR 04-OCT-2000; 2000US-0679272.
PR 28-MAR-2001; 2001US-0822827.
XX
PA (XUJ/) XU J.
PA (DILL/) DILLON D C.
PA (MITC/) MITCHAM J L.
PA (HARL/) HARLOCKER S L.
PA (JIAN/) JIANG Y.
PA (KALO/) KALOS M D.
PA (FANG/) FANGER G R.
PA (RETT/) RETTER M W.
PA (STOL/) STOLK J A.
PA (DAYC/) DAY C H.
PA (VEDV/) VEDVICK T S.
PA (CART/) CARTER D.
PA (LISX/) LI S X.
PA (WANG/) WANG A.
PA (SKET/) SKETKY Y A W.
PA (HEPL/) HEPLER W T.
PA (HEND/) HENDERSON R A.
PA (HURA/) HURAL J.
PA (MCNE/) MCNEILL P D.
PA (HOUG/) HOUGHTON R L.
PA (DBAS/) Y DE BASSOLS C V.
PA (FOYT/) FOY T M.
XX
XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD,
PI Fanger GR, Retter MM, Stolk JA, Day CH, Vedvick TS, Carter D;
PI Li SX, Wang A, Skelky YAM, Hepler WT, Henderson RA, Hural J;
PI Mcneill PD, Houghton RL, Y De Bassols CV, Foy TM;

XX
DR WPI; 2003-352711/33.
XX
PT New fusion protein comprising prostate-specific polypeptides, or its
PT immunogenic portions, useful for diagnosing, preventing and/or treating
PT cancer, particularly prostate cancer
XX
PS Example 2; SEQ ID NO 932; 85pp; English.
XX
CC The invention describes a fusion protein comprising at least one amino
CC acid sequence of immunogenic portions of any of the 3 sequences not
CC defined in the specification, or sequences having at least 70 or 90 %
CC sequence identity to any one of the 35 sequences defined in the USPTO
CC web site, which is encoded by any of the 4 nucleotide sequences not
CC defined in the specification. The fusion protein, composition and
CC methods are useful for diagnosing, preventing and/or treating cancer,
CC particularly prostate cancer. The proteins are useful as markers to
CC indicate the presence or absence of cancer. This is the amino acid
CC sequence of a prostate cancer therapy associated protein.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from the US patent office at
CC seqdata.uspto.gov/sequence.html?DocID=US20020192763.
XX
SQ Sequence 492 AA;
Query Match 99.8%; Score 1540; DB 24; Length 492;
Best Local Similarity 99.6%; Pred. No. 9.2e-136;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MKLNTSAGNVDIYKLYHSDACSSKAVSLRCLACGVNINSSRSRIVGSESALPGAMPW 60
Db 210 MKLNTSAGNVDIYKLYHSDACSSKAVSLRCLACGVNINSSRSRIVGSESALPGAMPW 269
Qy 61 QVSLHVQNVHVCSSGSIITPEMIVTAHCEKPLNPMHTAFAGILROSFMFYGAGYQVE 120
Db 270 QVSLHVQNVHVCSSGSIITPEMIVTAHCEKPLNPMHTAFAGILROSFMFYGAGYQVE 329
Qy 121 KVISHPNVDSTKKNNDIALMKLOKPLTFNDLVKPVCLPNEGMLQPEQLCMISGMGATEE 180
Db 330 KVISHPNVDSTKKNNDIALMKLOKPLTFNDLVKPVCLPNEGMLQPEQLCMISGMGATEE 389
Qy 181 KGMTSEVLNAAKVLLIETORCNSRYVDNLITPAMI CAGFLQGNVDS CGDSCGXLVTSK 240
Db 330 KGMTSEVLNAAKVLLIETORCNSRYVDNLITPAMI CAGFLQGNVDS CGDSCGXLVTSK 449
Qy 241 NNIMWLIGDTSWGSCKAKAYRPGVGNVVFDTMIYRORADG 283
Db 450 NNIMWLIGDTSWGSCKAKAYRPGVGNVVFDTMIYRORADG 492

Search completed: August 1, 2003, 18:11:40
Job time : 46 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2003, 18:09:03 ; Search time 11 Seconds

(without alignments)
1209.869 Million cell updates/sec

Title: US-09-988-975a-1

Sequence: 1 MKLNTSAGNVDIYKKLYHSD.....VYGNVWVFTDWIRQRADG 283

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------|
| 1 | 1540 | 99.8 | 492 | 1 | TMS2_HUMAN |
| 2 | 1264.5 | 82.0 | 490 | 1 | TMS2_MOUSE |
| 3 | 694 | 45.0 | 453 | 1 | TMS3_MOUSE |
| 4 | 686.5 | 44.5 | 454 | 1 | TMS3_HUMAN |
| 5 | 592.5 | 38.4 | 457 | 1 | TMS5_HUMAN |
| 6 | 575.5 | 37.3 | 455 | 1 | TMS5_MOUSE |
| 7 | 538.5 | 36.2 | 437 | 1 | TMS4_HUMAN |
| 8 | 555.5 | 36.0 | 638 | 1 | KAL_HUMAN |
| 9 | 541 | 35.1 | 638 | 1 | KAL_RAT |
| 10 | 536.5 | 34.8 | 1035 | 1 | ENTK_BOVIN |
| 11 | 528.5 | 34.3 | 1019 | 1 | ENTK_HUMAN |
| 12 | 528.5 | 34.1 | 638 | 1 | KAL_MOUSE |
| 13 | 520.5 | 33.7 | 417 | 1 | HEPS_HUMAN |
| 14 | 520 | 33.7 | 625 | 1 | FALL_HUMAN |
| 15 | 518.5 | 33.6 | 418 | 1 | HATT_HUMAN |
| 16 | 518.5 | 33.6 | 855 | 1 | STI4_MOUSE |
| 17 | 513.5 | 33.3 | 1069 | 1 | ENTK_MOUSE |
| 18 | 512.5 | 33.2 | 855 | 1 | STI4_HUMAN |
| 19 | 504.5 | 32.7 | 812 | 1 | PLMN_MOUSE |
| 20 | 503.5 | 32.6 | 416 | 1 | HEPS_RAT |
| 21 | 501.5 | 32.6 | 1034 | 1 | ENTK_PIG |
| 22 | 501.5 | 32.5 | 436 | 1 | HEPS_MOUSE |
| 23 | 492 | 31.9 | 421 | 1 | ACRO_HUMAN |
| 24 | 491.5 | 31.9 | 324 | 1 | TEST_MOUSE |
| 25 | 490 | 31.8 | 1042 | 1 | CORI_HUMAN |
| 26 | 486.5 | 31.5 | 422 | 1 | DESI_HUMAN |
| 27 | 481.5 | 31.2 | 415 | 1 | ACRO_PIG |
| 28 | 480.5 | 31.1 | 436 | 1 | ACRO_MOUSE |
| 29 | 466.5 | 30.2 | 431 | 1 | ACRO_RABBIT |
| 30 | 462 | 29.9 | 290 | 1 | MPN_HUMAN |
| 31 | 461.5 | 29.9 | 437 | 1 | ACRO_RAT |
| 32 | 460.5 | 29.8 | 311 | 1 | TRYG_MOUSE |
| 33 | 459 | 29.7 | 1113 | 1 | CORI_MOUSE |

| | | | | | | |
|----|-------|------|-----|---|------------|--------------------|
| 34 | 458.5 | 29.7 | 321 | 1 | TRYG_HUMAN | Q9ur22 homo sapien |
| 35 | 458.5 | 29.7 | 790 | 1 | PLMN_PIG | P06867 sus scrofa |
| 36 | 454 | 29.4 | 273 | 1 | TRYG_SHEEP | Q9x822 ovis aries |
| 37 | 454 | 29.4 | 810 | 1 | PLMN_BURBU | Q93483 etiacus e |
| 38 | 453 | 29.4 | 317 | 1 | BSS4_HUMAN | Q9g214 homo sapien |
| 39 | 452 | 29.3 | 343 | 1 | BSS4_MOUSE | Q16651 homo sapien |
| 40 | 451.5 | 29.3 | 306 | 1 | BSS4_MOUSE | Q9e10 mus musculu |
| 41 | 450.5 | 29.2 | 314 | 1 | TEST_HUMAN | Q9e60 homo sapien |
| 42 | 450 | 29.2 | 342 | 1 | BSS8_RAT | Q9e87 rattus norv |
| 43 | 449.5 | 29.1 | 810 | 1 | PLMN_MACMU | P12545 macaca mula |
| 44 | 447.5 | 29.0 | 343 | 1 | PLMN_SHEEP | P61286 ovis aries |
| 45 | 444 | 28.8 | 812 | 1 | PLMN_BOVIN | P06868 bos taurus |

ALIGNMENTS

RESULT 1
TMS2_HUMAN STANDARD; PRT; 492 AA.
AC O15353; Q9EXX1; 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DB Transmembrane protease, serine 2 precursor (EC 3.4.21.-).
GN TMPSR2 OR PRSS10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9746814; PubMed=9335052;
RA Paoloni-Giacobino A., Chen H., Peltsch M.C., Rosier C.,
RA Antonarakis S.B.;
RT Cloning of the TMPSR2 gene, which encodes a novel serine protease
RT with transmembrane, LDLR, and SRCR domains and maps to 21q22.3.";
RL Genomics 44:309-320 (1997). **Sp+ 97**
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21309069; PubMed=11414763;
RA Teng D.H., Chen Y., Lian L., He P.C., Tavrilian S.V., Mong A.K.;
RT Mutation analyses of 268 candidate genes in human tumor cell lines.";
RL Genomics 74:352-364 (2001).
RN [3]
RP SEQUENCE FROM N.A., AND MUTAGENESIS.
RX MEDLINE=21139112; PubMed=11245484;
RA Afar D.B.H., Vivanco I., Hubert R.S., Kuo J., Chen B., Saffran D.C.,
RA Raitano A.B., Jakobovits A.;
RT Catalytic cleavage of the androgen-regulated TMPSR2 protease results
RT in its secretion by prostate and prostate cancer epithelia.";
RL Cancer Res. 61:1686-1692 (2001).
RN [4]
RP TISSUE SPECIFICITY.
RX MEDLINE=21104370; PubMed=11169526;
RA Vaerla W.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Vihko P.T.;
RT "Expression of transmembrane serine protease TMPSR2 in mouse and
RT human tissues.";
RL J. Pathol. 193:134-140 (2001).
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ACTIVATED BY
CC CLEAVAGE AND SECRETED.
CC -1- TISSUE SPECIFICITY: EXPRESSED STRONGLY IN SMALL INTESTINE. ALSO
CC EXPRESSED IN PROSTATE, COLON, STOMACH, AND SALIVARY GLAND.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: Contains 1 LDL-receptor class A domain.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

```

CC or send an email to license@ib-stb.ch).
CC -----
DR EMBL; U75329; AAC51784.1; -.
DR EMBL; AF123453; AAD37117.1; -.
DR EMBL; AF270487; AAK29280.1; -.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.247; -.
DR Genew; HGNC:11876; TMRSS2.
DR MIM; 602060; -.
DR GO; GO:0005867; C:Integral to plasma membrane; TAS.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR001190; Strc_receptor.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLA; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00202; Tryp_Spc; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS50068; LDLRA_2; 1.
DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen;
KW Polymorphism.
FT CHAIN 1 255 TRANSMEMBRANE PROTEASE, SERINE 2, NON-
FT CHAIN 1 255 CATALYTIC CHAIN.
FT CHAIN 256 492 TRANSMEMBRANE PROTEASE, SERINE 2,
FT CHAIN 256 492 CATALYTIC CHAIN.
FT DOMAIN 1 84 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 85 105 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT DOMAIN 106 492 (POTENTIAL).
FT DOMAIN 112 149 LDL-RECEPTOR CLASS A.
FT DOMAIN 150 242 SRCR.
FT DOMAIN 256 492 SERINE PROTEASE.
FT ACT_SITE 296 345 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 345 345 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 441 441 CHARGE RELAY SYSTEM.
FT SITE 255 256 CLEAVAGE (POTENTIAL).
FT DISULFID 113 126 BY SIMILARITY.
FT DISULFID 120 139 BY SIMILARITY.
FT DISULFID 133 148 BY SIMILARITY.
FT DISULFID 172 231 BY SIMILARITY.
FT DISULFID 185 241 BY SIMILARITY.
FT DISULFID 244 365 INTERCHAIN (BY SIMILARITY).
FT DISULFID 281 297 BY SIMILARITY.
FT DISULFID 410 426 BY SIMILARITY.
FT DISULFID 437 443 BY SIMILARITY.
FT CARBOHYD 213 213 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 449 449 K->N (IN DBSNP:1056602).
FT MUTAGEN 255 255 /FTID=VAR 011692.
FT MUTAGEN 441 441 R->Q: LOSS OF CLEAVAGE.
FT CONFLICT 160 160 S->A: LOSS OF ACTIVITY.
FT CONFLICT 242 242 M->V (IN REF. 3).
FT CONFLICT 329 329 I->L (IN REF. 1).
FT CONFLICT 329 329 E->Q (IN REF. 1).
FT CONFLICT 489 491 RMD->KAN (IN REF. 1).
SQ SEQUENCE 492 AA; 53891 MW; 53891 MD; CAB44FD17A9076B CRC64;

Query Match 99.8%; Score 1540; DB 1; Length 492;
Best Local Similarity 99.6%; Pred No. 1.5e-131;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Db 270 QVSLHVNQVHVGCGSITTPETVITNAHCVEKPLNPNMHTAFAGILRSGFMRYGAGYQVE 329
Qy 121 KVISHPNYSKTKKNDIALMELQKELTENDLVKPVCLPDPGMDQPEOLCWSIGMGATEE 180
Db 330 KVISHPNYSKTKKNDIALMELQKELTENDLVKPVCLPDPGMDQPEOLCWSIGMGATEE 389
Qy 181 KAKTSEVITNAKAVLLIFORCKSRVYDNLITPAMICAGFLOGANDSCGDSGGLVTSK 240
Db 390 KAKTSEVITNAKAVLLIFORCKSRVYDNLITPAMICAGFLOGANDSCGDSGGLVTSK 449
Qy 241 NNIMWLIGDTWGSAGCAKARYGVYGNVVFPTDWTYRQBADG 283
Db 450 NNIMWLIGDTWGSAGCAKARYGVYGNVVFPTDWTYRQBADG 492

RESULT 2
TMS2 MOUSE STANDARD; PRT; 490 AA.
AC Q9UJ08; Q9UKC4; Q9Y82;
DT 16-OCT-2001 (Rel. 40; Created)
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DT 28-FEB-2003 (Rel. 41; Last annotation update)
DE Transmembrane protease, serine 2 (EC 3.4.21.-) (Epitheliasin) (Plasmin
DE Transmembrane protein X).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC STRAIN=BALB/c;
RX MEDLINE=21104370; PubMed=11169526;
RA Vaarala M.H., Potvari K.S., Kellokumpu S., Kyllonen A.P., Vinko P.T.;
RT "Expression of transmembrane serine protease TMRSS2 in mouse and
RT human tissues.";
RL J. Pathol. 193:134-140 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Han J., Kim S.;
RT "Putative transmembrane protease X.";
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN=BALB/c;
RA Jacquinet B.J., Rao N.V., Rao G.N., Hoidal J.R.;
RT "A novel mosaic serine protease, epitheliasin.";
RL Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ACTIVATED BY
CC CLEAVAGE AND SECRETED (By similarity).
CC -1- TISSUE SPECIFICITY: EXPRESSED MAINLY IN PROSTATE AND KIDNEY.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC -1- SIMILARITY: Contains 1 LDL-receptor class A domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-stb.ch/announce/
CC or send an email to license@ib-stb.ch).
CC -----
DR EMBL; AF199362; AAF97867.1; -.
DR EMBL; AF243500; AAF64186.1; -.
DR EMBL; AF113596; AAF21308.1; -.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.247; -.
DR MGD; MGI:1354381; Tmrps2.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Ser_protease_Try.

```

```

DR InterPro: IPR001190; Srcr_receptor.
DR Pfam: PF00057; ldl_recept_a; 1.
DR Pfam: PF00530; SRCR; 1.
DR Pfam: PF00089; tryptsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00192; LDLA; 1.
DR SMART: SM00202; SR; 1.
DR SMART: SM00202; TRYP_SPC; 1.
DR PROSITE: PS01209; LDLRA_1; 1.
DR PROSITE: PS50068; LDLRA_2; 1.
DR PROSITE: PS00420; SRCR_1; FALSE_NEG.
DR PROSITE: PS50287; SRCR_2; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen.
FT CHAIN 1 253
FT TRANSMEMBRANE PROTEASE, SERINE 2, NON-
FT CATALYTIC CHAIN.
FT CHAIN 254 490
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT DOMAIN 1 83
FT TRANSMEM 84 104
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT EXTRACELLULAR (POTENTIAL).
FT LDL-RECEPTOR CLASS A.
FT SRCR.
FT SERINE PROTEASE.
FT CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 254 490
FT ACT_SITE 324 490
FT ACT_SITE 333 439
FT ACT_SITE 439 439
FT SITE 253 254
FT DISULFID 76 125
FT DISULFID 119 125
FT DISULFID 132 125
FT DISULFID 171 230
FT DISULFID 184 240
FT DISULFID 243 363
FT DISULFID 279 295
FT DISULFID 408 424
FT DISULFID 435 463
FT CARBOHYD 111 111
FT CARBOHYD 212 474
FT CARBOHYD 474 474
FT CONFLICT 122 122
FT CONFLICT 178 178
FT CONFLICT 320 320
FT CONFLICT 474 474
FT SEQUENCE 490 AA; 53479 MW; 07D2B03EAD8A1A9 CRC64;
SQ

```

```

Query Match 82.0%; Score 1264.5; DB 1; Length 490;
Best Local Similarity 82.3%; Pred. No. 1.1e-106;
Matches 232; Conservative 20; Mismatches 29; Indels 1; Gaps 1;

```

```

QY 1 MKLNTSAGNDVIYKLYHSDACSSKAVNGLRCIACSVNINSSQSRHIVGGESELPQAMPW 60
DB 209 MKLNTSAGNDVIYKLYHSDACSSKAVNGLRCIACSVNINSSQSRHIVGGESELPQAMPW 267
QY 61 QVSLHVNQVHVGGSIIITPEWITVAHCEYKPLANNPMTAAGILROSEFMFYAGYQVE 120
DB 268 QVSLHVNQVHVGGSIIITPEWITVAHCEYKPLANNPMTAAGILROSEFMFYAGYQVE 327
QY 121 KVISHPNDSTKRNNDIALMKLQKPLTPNDLYKPVCLPNPQMLDPEQLCWISGWTATEE 180
DB 328 KVISHPNDSTKRNNDIALMKLQKPLTPNDLYKPVCLPNPQMLDPEQLCWISGWTATEE 387
QY 181 KGTSEVNTAAKAVLLIETORCNSTRYYVDLITPAMCAEFLGQVDSGGSGGXLYTSK 240
DB 388 KGTSEVNTAAKAVLLIETORCNSTRYYVDLITPAMCAEFLGQVDSGGSGGXLYTSK 447
QY 241 NNIMWLIDTWSGSCARAPRPGYGNVWFTDWIYRQNRAD 282
DB 448 NNIMWLIDTWSGSCARAPRPGYGNVWFTDWIYRQNRAD 489

```

```

RESULT 3
ID TMS3_MOUSE STANDARD; PRT; 453 AA.
AC O8K1T0; Q8VDE0;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Transmembrane protease, serine 3 (EC 3.4.21.-).
GN TMRSS3.
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID:10090;
RN [1]
RP SEQUENCE FROM N.A., SUBCELLULAR LOCATION, AND FUNCTION IN ENAC
RP CLEAVAGE.
RX PubMed:12393794;
RA Guipponi M., Vassilaux G., Wattenhofer M., Shibuya K., Vazquez M.,
RA Dougherty L., Scamuffe N., Guida E., Okui M., Rossier C., Hancock M.,
RA Buchet K., Raymond A., Hummler B., Marzella P.L., Kudoh J.,
RA Shimizu N., Scott H.S., Antonarakis S.E., Rossier B.C.;
RT "The transmembrane serine protease (TMRSS3) mutated in deafness
RT DFNB8/10 activates the epithelial sodium channel (ENaC) in vitro.";
RL Hum. Mol. Genet. 11:2829-2836(2002).
CC -1- FUNCTION: Probable protease. Seems to be capable of activating
CC ENaC.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic
CC reticulum.
CC -1- TISSUE SPECIFICITY: Expressed in the spiral ganglion, the cells
CC supporting the organ of Corti and the stria vascularis.
CC -1- PTM: Undergoes autoproteolytic activation.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC -1- SIMILARITY: Contains 1 LDL-receptor class A domain.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announce/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; AJ429216; CAD23137.1; -.
DR EMBL; AJ300738; CAC83350.1; -.
DR HSSP; P00761; IAN1.
DR MGD; MGI:2155445; Tmrps3.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR002172; LDL_receptor_A.
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR001190; Srcr_receptor.
DR Pfam: PF00057; ldl_recept_a; 1.
DR Pfam: PF00089; tryptsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00192; LDLA; 1.
DR SMART: SM00202; SR; 1.
DR SMART: SM00202; TRYP_SPC; 1.
DR PROSITE: PS01209; LDLRA_1; 1.
DR PROSITE: PS50068; LDLRA_2; 1.
DR PROSITE: PS50287; SRCR_2; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen;
KW Endoplasmic reticulum.
FT DOMAIN 1 48
FT TRANSMEM 49 69
FT TRANSMEM 70 453
FT DOMAIN 72 108
FT DOMAIN 104 205
FT DOMAIN 217 448
FT SIGNAL-ANCHOR (POTENTIAL).
FT CYTOPLASMIC (POTENTIAL).
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT EXTRACELLULAR (POTENTIAL).
FT LDL-RECEPTOR CLASS A.
FT SRCR.
FT SERINE PROTEASE.

```

FT ACT_SITE 257 257 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 304 304 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 400 400 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT SITE 216 217 CLEAVAGE (POTENTIAL).
 FT DISULFID 73 85 BY SIMILARITY.
 FT DISULFID 79 98 BY SIMILARITY.
 FT DISULFID 92 107 BY SIMILARITY.
 FT DISULFID 129 194 BY SIMILARITY.
 FT DISULFID 142 204 BY SIMILARITY.
 FT DISULFID 207 324 BY SIMILARITY.
 FT DISULFID 242 258 BY SIMILARITY.
 FT DISULFID 338 406 BY SIMILARITY.
 FT DISULFID 369 385 BY SIMILARITY.
 FT DISULFID 424 424 BY SIMILARITY.
 FT CARBOHYD 221 221 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 117 117 L -> H (IN REF. 1; CAC83350).
 FT CONFLICT 246 246 V -> I (IN REF. 1; CAC83350).
 SQ SEQUENCE 453 AA; 49491 MM; 1ABC8P10AF6E1EF6 CRC64;

Query Match 45.0%; Score 694; DB 1; Length 453;
 Best Local Similarity 49.6%; Pred. No. 3.4e-55;
 Matches 135; Conservative 39; Mismatches 92; Indels 6; Gaps 4;

QY 12 IYKRLVSDACSKAVVSLRCLACGVNINSSRSHVGGESALPGAMPQVSLHYQNVHV 71
 DB 184 LHSVYMRGCTSGHVTLKSCACSTRGYG--PRIVGGNNSLTQMPQVSLQPGYHL 241
 -QY 72 CGGSIITPEWITVAHAQVEKPLNPMHTAFAGILROSEMFYGA-GYQVEKVIHPNDYD 130
 DB 242 CGGSVITPEWITVAHAQV-YDLVHKSMVTQVGLV--SLMDSVPVSHLVKELIYHSKRP 298
 QY 131 KTKRNDIALMKLQKPLTNDLVKPYCLNPGMMLQPEOLCWISGMGATEEKGKTSVZANA 190
 DB 299 KRLNDIALMKLSEPLTDETRIQPLCPNSENPFDPGKLCWTSGMGATEDGDAAPVLNH 358
 QY 191 AKVLLIETQRNSRVYVNLITPAMICAGFLQGVNDSGCGSGXGLVYSKNNIMLIGDT 250
 DB 359 AAVPLISNKICNHRVYGGIISPSMLCAGYLKGVNDSGCGSGGPLVCCERRLMKLVGAT 418
 QY 251 SMGSGCAKAVRPGVYGVNMTFTMYRQMRAD 282
 DB 419 SFGIGCAEVNKGYYTRITSPIDWTHQLEERD 450

RESULT 4
 TMS3 HUMAN
 ID TMS3 HUMAN STANDARD; PRT; 454 AA.
 AC P57727;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Transmembrane protease, serine 3 (BC 3.4.21.-) (Serine protease
 TADG-12) (Tumor associated differentially-expressed gene-12 protein).
 GN TMPSRS3 OR TADG12 OR ECHOS1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS A AND T).
 RC TISSUE=Ovarian carcinoma;
 RX MEDLINE=20521358; PubMed=11068177;
 RA Underwood L.J., Shigemasa K., Tanimoto H., Beard J.B., Schneider E.N.,
 RA Wang Y., Parmley T.H., O'Brien T.J.;
 RT "Ovarian tumor cells express a novel multi-domain cell surface serine
 RT protease";
 RL Biochim. Biophys. Acta 1502:337-350(2000).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS A AND D) AND VARIANT ILB-53.
 RX MEDLINE=20578749; PubMed=11137999;
 RA Scott H.S., Kudoh J., Wattenhofer M., Shibuya K., Berry A., Chrast R.,
 RA Guipponi M., Wang J., Kawasaki K., Asakawa S., Minoshima S.,
 RA Youngs F., Mehdi S.O., Radhakrishna U., Papasavvas M.P., Gehrig C.,

RA Rosier C., Korostilchewsky M., Gal A., Shimizu N., Bonne-Tamir B.,
 RA Antonarakis S.B.;
 RT "Insertion of beta-satellite repeats identifies a transmembrane
 RT protease causing both congenital and childhood onset autosomal
 RT recessive deafness";
 RL Nat. Genet. 27:59-63(2001).
 RN [3]
 RP SUBCELLULAR LOCATION, AND FUNCTION IN ENAC CLEAVAGE.
 RX PubMed=12393794;
 RA Guipponi M., Vuagniaux G., Wattenhofer M., Shibuya K., Vazquez M.,
 RA Dougherty L., Scamuffa N., Guida E., Okui M., Rosier C., Hancock M.,
 RA Bucher K., Raymond A., Hummler B., Marzella P.L., Kudoh J.,
 RA Shimizu N., Scott H.S., Antonarakis S.B., Rosier B.C.;
 RT "The transmembrane serine protease (TMPSRS3) mutated in deafness
 RT DFN8/10 activates the epithelial sodium channel (ENaC) in vitro";
 RL Hum. Mol. Genet. 11:2829-2836(2002).
 RN [4]
 RP VARIANTS DFN8/DFNB10 CYS-251 AND LEU-404.
 RX MEDLINE=21354482; PubMed=11462234;
 RA Mamoudi S., Antonarakis S.B., Schwede T., Ghorbel A.M., Gratri M.,
 RA Papasavvas M.P., Ditra M., Elgaled-Bouilla A., Wattenhofer M.,
 RA Rosier C., Scott H.S., Ayadi H., Guipponi M.;
 RT "Novel missense mutations of TMPSRS3 in two consanguineous Tunisian
 RT families with non-syndromic autosomal recessive deafness";
 RL Hum. Mutat. 18:101-108(2001).
 RN [5]
 RP VARIANTS DFN8/DFNB10 TRP-109, PHE-194 AND ARG-407, AND VARIANTS
 RP ILB-53; SER-111 AND VAL-253.
 RX MEDLINE=21317610; PubMed=11424922;
 RA Ben-Yosef T., Wattenhofer M., Riazuddin S., Ahmed Z.M., Scott H.S.,
 RA Kudoh J., Shibuya K., Antonarakis S.B., Bonne-Tamir B.,
 RA Radhakrishna U., Naz S., Ahmed Z., Riazuddin S., Pandya A.,
 RA Nance W.E., Wilcox E.R., Friedman T.B., Morell R.J.;
 RT "Novel mutations of TMPSRS3 in four DFN8/810 families segregating
 RT congenital autosomal recessive deafness";
 RL J. Med. Genet. 38:396-400(2001).
 RN [6]
 RP VARIANT DFN8/DFNB10 GLY-103, AND VARIANTS ASN-173 AND THR-426.
 RX MEDLINE=21904597; PubMed=11907649;
 RA Wattenhofer M., Di Iorio V., Rablione R., Dougherty L., Pampas A.,
 RA Schwede T., Montserrat-Sentis B., Arbones L., Illades T.,
 RA Paegleabiscaglia A., D'Amelio M., Alwan S., Rosier C., Dahl H.-H.M.,
 RA Petersen M.B., Bectivill X., Gasparini P., Scott H.S.,
 RA Antonarakis S.B.;
 RT "Mutations in the TMPSRS3 gene are a rare cause of childhood
 RT non-syndromic deafness in Caucasian patients";
 RL J. Mol. Med. 80:124-131(2002).
 CC -1- FUNCTION: Probable protease. Seems to be capable of activating
 CC ENaC.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic
 CC reticulum.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=4;
 CC Name=A;
 CC isoId=P57727-1; Sequence=Displayed;
 CC Name=B; Synonyms=C;
 CC isoId=P57727-2; Sequence=VSP_005391;
 CC Name=D;
 CC isoId=P57727-3; Sequence=VSP_005392;
 CC Name=E; Synonyms=F;
 CC isoId=P57727-4; Sequence=VSP_005393; VSP_005394;
 CC -1- TISSUE SPECIFICITY: Expressed in many tissues including fetal
 CC cochlea. Isoform T is found at increased levels in some
 CC carcinomas.
 CC -1- PTM: Undergoes autophosphorylation.
 CC -1- DISEASE: Defects in TMPSRS3 are a cause of childhood-onset
 CC autosomal recessive neurosensory deafness 8 (DFNB8) [MIM:601072].
 CC -1- DISEASE: Defects in TMPSRS3 are a cause of congenital autosomal
 CC recessive neurosensory deafness 10 (DFNB10) [MIM:605316].
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 CC -1- SIMILARITY: Contains 1 SRCR domain.
 CC -1- SIMILARITY: Contains 1 LDL-receptor class A domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).

DR EMBL; AF201380; AAG37012.1; -
 DR EMBL; AB038157; BAB20077.1; -
 DR EMBL; AB038158; BAB20078.1; -
 DR EMBL; AB038159; BAB20079.1; -
 DR EMBL; AB038160; BAB20080.1; -
 DR HSSP; P00763; IDPO.
 DR MEROPS; S01.079; -
 DR Genew; HGNC:11877; TMPRSS3.
 DR MIM; 605511; -
 DR MIM; 605316; -
 DR GO; GO:0016021; C:integral to membrane; NAS.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; NAS.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR Pfam; PF00057; 1d1_recept_a; 1.
 DR Pfam; PF00089; tryptsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00192; LDLa; 1.
 DR SMART; SM00202; SR; 1.
 DR SMART; SM00202; TYP_Spc; 1.
 DR PROSITE; PS01209; LDLRA_1; 1.
 DR PROSITE; PS0068; LDLRA_2; 1.
 DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
 DR PROSITE; PS0240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_SER; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KM Hydroxylase; Serine protease; Transmembrane; Signal-anchor; Zymogen;
 KM Endoplasmic reticulum; Deafness; Alternative splicing;
 KM Disease mutation; Polymorphism.
 FT DOMAIN 1 48
 FT TRANSMEM 49 69
 FT DOMAIN 70 454
 FT DOMAIN 72 108
 FT DOMAIN 109 205
 FT DOMAIN 217 454
 FT ACT_SITE 257 304
 FT ACT_SITE 304 304
 FT ACT_SITE 401 401
 FT SITE 216 217
 FT DISULFID 73 85
 FT DISULFID 79 98
 FT DISULFID 92 107
 FT DISULFID 129 194
 FT DISULFID 142 204
 FT DISULFID 207 324
 FT DISULFID 242 258
 FT DISULFID 338 407
 FT DISULFID 370 386
 FT DISULFID 397 425
 FT CARBOHYD 221 221
 FT VARSPIC 1 127
 FT VARSPIC 318 454
 FT VARSPIC 261 293

VARSPIC 261 293

FT FT
 FT VARSPIC 294 454
 FT FT
 FT VARIANT 53 53
 FT FT
 FT VARIANT 103 103
 FT FT
 FT VARIANT 109 109
 FT FT
 FT VARIANT 111 111
 FT FT
 FT VARIANT 173 173
 FT FT
 FT VARIANT 194 194
 FT FT
 FT VARIANT 251 251
 FT FT
 FT VARIANT 253 253
 FT FT
 FT VARIANT 404 404
 FT FT
 FT VARIANT 407 407
 FT FT
 FT VARIANT 426 426
 FT FT
 FT CONFLICT 46 54

Query Match 44.5%; Score 686.5; DB 1; Length 454;
 Best Local Similarity 50.2%; Pred. No. 1,66-54;
 Matches 137; Conservative 37; Mismatches 92; Indels 7; Gaps 5;

QY 12 IYKLYHSAGCSKAVSISACIACGVNANSSQSRIVGESALPGAMPQVSLYQNVAV 71
 DB 184 IHRSVYVEGCASGVVTLQCTACGHRGYS--SRIVGNMSELISQMPQASIQFGGYH 241
 QY 72 CGGSITPEWITAAHCVEKPLANNPMHTAFAGILRQSPMFGA-GYVEKYSHPNVYS 130
 DB 242 CGGSVITPMITTAHCV-YDLYKRSWITQVLV-SLDNPAASHLVEKYSHPNVYS 298
 QY 131 KTKNDIALMLQKPLTFNDLVKVCPLPFGMMLQPEQLCWTSGWGTRE-KGKTSEVLA 189
 DB 299 KRLGNDIALMLQAGLTFNEMIPVCLRNSENPFDGKVCWTSGWGTREKGDGADSPVLN 358
 QY 190 AAKVLLITQRCNSRYVDNLITPMLICAGFLQGVNDSQGDGGLVTSKNNIMVLID 249
 DB 359 HAAVPLINIKICNHRDVGGLISPSMLCAGLYTGVDSCQDGGGFLVCGERRLMTLVGA 418
 QY 250 TSMGSCAKAYRPGYGVNMTPTDWTYROMRD 282
 DB 419 TSFGIGCARVNRPGVYTVTSFLDWHOMERD 451

RESULT 5
 TMS5_HUMAN STANDARD; PRT; 457 AA.
 ID TMS5_HUMAN
 AC Q9H353;
 DT 16-OCT-2001 (Rel. 40; Created)
 DT 16-OCT-2001 (Rel. 40; Last sequence update)
 DT 28-FEB-2003 (Rel. 41; Last annotation update)
 DE Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinein).
 GN TMPRSS3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX PubMed=11741986;
 RA Yamaguchi N., Okui A., Yamada T., Nakazato H., Mitsui S.;
 RT "Spinein/TMPRSS3, a novel transmembrane serine protease, cloned from
 human spinal cord.";

```

RL J. Biol. Chem. 277:6806-6812(2002).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
CC -1- TISSUE SPECIFICITY: Brain-specific. Predominantly expressed in
CC neurons, in their axons, and at the synapses of motoneurons in the
CC spinal cord.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: Contains 1 SRCR domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
-----
CC EMBL: AB028140; BAB20375.1; -.
CC HSSP: P00763; IDPO.
CC Genew: HGNC:14908; TMPS55.
CC MIM: 606751; -.
CC MEROPS: S01.313; -.
CC InterPro: IPR001314; Chymotrypsin.
CC InterPro: IPR001254; Ser. protease Try.
CC InterPro: IPR001190; Ser. receptor.
CC Pfam: PF00089; trypsin_1.
CC PRINTS: PR00722; CHYMOTRYPSIN.
CC SMART: SM00020; TRYD_Spc; 1.
CC PROSITE: PS50240; TRYPSIN_DOM; 1.
CC PROSITE: PS00134; TRYPSIN_HIS; 1.
CC PROSITE: PS00135; TRYPSIN_SER; 1.
CC PROSITE: PS00420; SRCR_1; FALSE_NEG.
CC PROSITE: PS50287; SRCR_2; FALSE_NEG.
CC KEGG: Hydrolase; Serine protease; Transmembrane; Signal-anchor;
CC Glycoprotein.
CC TRANSMEM 1 49 CYTOPLASMIC (POTENTIAL).
CC DOMAIN 50 70 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC (POTENTIAL).
CC EXTRACELLULAR (POTENTIAL).
CC DOMAIN 71 457 SRCR.
CC DOMAIN 112 207 SERINE PROTEASE.
CC ACT_SITE 218 457 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC ACT_SITE 258 258 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC ACT_SITE 308 308 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC ACT_SITE 405 405 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC SITE 217 218 CLEAVAGE (POTENTIAL).
CC DISULFID 135 196 BY SIMILARITY.
CC DISULFID 148 206 BY SIMILARITY.
CC DISULFID 209 328 BY SIMILARITY.
CC DISULFID 243 259 BY SIMILARITY.
CC DISULFID 374 390 BY SIMILARITY.
CC DISULFID 401 429 BY SIMILARITY.
CC CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 195 195 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 319 319 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 457 AA; 49574 MW; 64406AB495A2651 CRC64;
Query Match 38.4%; Score 592.5; DB 1; Length 457;
Best Local Similarity 42.3%; Pred. No. 5e-46;
Matches 121; Conservative 38; Mismatches 114; Indels 13; Gaps 5;

```

```

QY 172 ISMGATESEKX-TSEVNAKAVLLIETORCNSRYVYVNLTPAMICAGFIQGVDSQCG 230
DB 344 VSGMHTPSTSTSDMLQDTVPPLSTOLCNSSCVISGALPPRLCAGYLDGRADACCG 403
QY 231 DSGGLVTSKNNIMWLIGDTSMGSCAXAPRGVGVNMTVPTDMY 276
DB 404 DSGGLVCPDDDTLRLVGVVSGRACAPRNPFGVYAXAFPLDWH 449
-----
RESULT 6
TMS5_MOUSE STANDARD: PRT; 455 AA.
ID TMS5_MOUSE
AC Q9ER04; Q9ER02; Q9ER03;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin).
GN TMPS55.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISORFMS 1; 2 AND 3).
RC TISSUE-Brain;
RA Mitsu S., Yamaguchi N.;
RT "cDNA cloning of mouse spinesin."
RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (ISORFMS 4).
RC TISSUE-Brain;
RA Mitsu S., Yamaguchi N.;
RT "Molecular cloning of mouse type 4 spinesin."
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=4;
CC Name=4;
CC IsoId=Q9ER04-1; Sequence=Displayed;
CC Name=1;
CC IsoId=Q9ER04-2; Sequence=VSP_005397, VSP_005398;
CC Name=2;
CC IsoId=Q9ER04-3; Sequence=VSP_005395;
CC Name=3;
CC IsoId=Q9ER04-4; Sequence=VSP_005396;
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: Contains 1 SRCR domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
-----
CC EMBL: AB016229; BAB20276.1; -.
CC EMBL: AB016230; BAB20277.1; -.
CC EMBL: AB016423; BAB20278.1; -.
CC EMBL: AB041037; BAB40328.1; -.
CC HSSP: P00763; IDPO.
CC MEROPS: S01.313; -.
CC MGD: MGI:1933407; TMPS55.
CC InterPro: IPR001254; Ser. protease Try.
CC InterPro: IPR001190; Ser. receptor.
CC Pfam: PF00089; trypsin_1.
CC SMART: SM00020; TRYD_Spc; 1.
CC PROSITE: PS50240; TRYPSIN_DOM; 1.
CC PROSITE: PS00134; TRYPSIN_HIS; 1.
CC PROSITE: PS00135; TRYPSIN_SER; 1.
CC PROSITE: PS00420; SRCR_1; FALSE_NEG.
CC PROSITE: PS50287; SRCR_2; 1.
CC Hydrolase; Serine protease; Transmembrane; Signal-anchor;

```



```

DR SMART; SMO0202; SR; 1.
DR PROSITE; PS01209; LDIRA_1; FALSE NEG.
DR PROSITE; PS50068; LDIRA_2; FALSE NEG.
DR PROSITE; PS00420; SRCR_1; FALSE NEG.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydroxylase; Serine protease; Transmembrane; Signal-anchor.
KM HYDROLASE; Serine protease; Transmembrane; Signal-anchor.
FT DOMAIN 1 38 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 39 59 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT DOMAIN 60 437 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 61 93 LDL-RECEPTOR CLASS A.1
FT DOMAIN 94 194 SRCR.
FT DOMAIN 205 437 SERINE PROTEASE.
FT ACT_SITE 245 245 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 290 290 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 387 387 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT SITE 204 205 CLEAVAGE (POTENTIAL).
FT DISULFID 64 83 BY SIMILARITY.
FT DISULFID 77 92 BY SIMILARITY.
FT DISULFID 127 183 BY SIMILARITY.
FT DISULFID 140 193 BY SIMILARITY.
FT DISULFID 196 310 BY SIMILARITY.
FT DISULFID 230 246 BY SIMILARITY.
FT DISULFID 356 372 BY SIMILARITY.
FT DISULFID 383 410 BY SIMILARITY.
FT CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1 31 M0DDPSDDPLNSLDVYKPKRKPRIMETTRK -> MSNPCA
FT SEQUENCE 437 AA; 48204 MW; 351B2FPDA657B12 CR64;
SQ NEVSPWRSES (IN REF. 2).
Query Match 36.2%; Score 558.5; DB 1; Length 437;
Best Local Similarity 43.9%; Pred. No. 5.6e-43;
Matches 118; Conservative 41; Mismatches 93; Indels 17; Gaps 8;
QY 19 SDACSKAVSVSLRCLACGVNLNNSRQSRIVGSGSALPGAMPQVSLAVQNVHVGSGSIIT 78
DB 180 SGPCLSGSLVSLHCLACGKSL---KTPRVVGGBEASVDSWMPQVSIQYQKHVGSGSILD 236
QY 79 PEMVITAAHCYKPLANPMHTAFACILR-QSFMFGAGYQYEVY---SHPRYSKTKQ 134
DB 237 PHWVITAAHCFKXH-TDVFNMTVRAGSDTGSF---PSLAVAKIILIFENPMY---PXD 288
QY 135 NDIALMKLOKPLTFNDIVVCLPAPNGMWLOPBOUCWISGKGTEBK-GKTSSEVNAAYV 193
DB 289 NDIALMKLOPPLTFSGTVAPICLPFPDEELTPATPLMIIGWGTYSKNGSMEDILQASV 348
QY 194 LLIEIQRNSRYYVNLITPAMIAGFLQGNVDSGCGSGGXLYVTSKNNIWMLIGDTSWG 253
DB 349 QVIDSTRCADADAYGGEVTERKMGAGIPRGVDTQCGSDSGPLMWGSDQ-WHVVGIVSVNG 407
QY 254 SGCAVAYRPGVYGNVFTDWIYRQMRAD 282
DB 408 YGCGGSPSTPGVYTKVSAVIMYIYMWKAR 436

```

```

OX NCBI_Taxid=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=86243359; PubMed=3521732;
RA Chung D.W., Fujikawa K., McMullen B.A., Davie B.W.;
RT "Human plasma prekallikrein, a zymogen to a serine protease that
RL contains four tandem repeats."
RL Biochemistry 25:2410-2417(1986).
RN (2)
RP SEQUENCE FROM N.A., AND VARIANTS SER-143; GLN-202 AND PRO-208.
RX MEDLINE=20487549; PubMed=11031105;
RA Yu H., Anderson P.J., Freedman B.I., Rich S.S., Bowden D.W.;
RT "Genomic structure of the human plasma prekallikrein gene;
RT identification of allelic variants, and analysis in end-stage renal
RT disease."
RL Genomics 69:225-234(2000).
RN (3)
RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX MEDLINE=91152016; PubMed=1998666;
RA McMullen B.A., Fujikawa K., Davie B.W.;
RT "Location of the disulfide bonds in human plasma prekallikrein: the
RT the molecule."
RL Biochemistry 30:2050-2056(1991).
CC -1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
CC -1- CATALYTIC ACTIVITY: Cleaves selectively Arg-Xaa and Lys-Xaa
CC bonds, including Lys-Arg and Arg-Ser bonds in (human)
CC kininogen to release bradykinin.
CC -1- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XIIA, WHICH CLEAVES
CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
CC -1- DISEASE: Defects in K1K1 are the cause of Fletcher factor
CC deficiency [MIM:229000]; a blood coagulation defect.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. PLASMA KALLIKREIN
CC SUBFAMILY.
CC -1- SIMILARITY: Contains 4 apple domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; M13143; AAA60153.1; -
DR EMBL; AF232742; AAF79940.1; -
DR EMBL; AF232734; AAF79940.1; JOINED.
DR EMBL; AF232735; AAF79940.1; JOINED.
DR EMBL; AF232736; AAF79940.1; JOINED.
DR EMBL; AF232737; AAF79940.1; JOINED.
DR EMBL; AF232738; AAF79940.1; JOINED.
DR EMBL; AF232739; AAF79940.1; JOINED.
DR EMBL; AF232740; AAF79940.1; JOINED.
DR EMBL; AF232741; AAF79940.1; JOINED.
DR PIR; A00921; K0HDP.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.212; -.
DR GeneW; HGNC:6371; KLKB1.
DR MIM; 229000; -.
DR GO; GO:0003807; F:plasma kallikrein activity; TAS.
DR GO; GO:000508; P:proteolysis and peptidolysis; TAS.
DR InterPro; IPR000177; Apple.
DR InterPro; IPR003014; PAN.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00024; PAN; 4.
DR Pfam; PF00089; trypsin; 1.

```



```

DR MEROPS; S01.212; -.
DR InterPro; IPR000177; Apple.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR003014; PAN.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00024; PAN; 4.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00005; APPLIEDOMAIN.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00223; APPLE; 4.
DR SMART; SM00202; Tryp_Spc; 1.
DR PROSITE; PS00495; APPLE; 4.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR HydroLase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
KM Fibrinolysis; Blood coagulation; Inflammatory response; Liver;
KM Repeat.
FT SIGNAL 1 19
FT CHAIN 20 390 PLASMA KALLIKREIN HEAVY CHAIN.
FT CHAIN 391 638 PLASMA KALLIKREIN LIGHT CHAIN.
FT DOMAIN 20 105 APPLE 1.
FT DOMAIN 110 195 APPLE 2.
FT DOMAIN 200 285 APPLE 3.
FT DOMAIN 291 376 APPLE 4.
FT DOMAIN 389 621 SERINE PROTEASE.
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 396 396 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 453 453 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .) (PROBABLE).
FT ACT_SITE 434 434 CHARGE RELAY SYSTEM.
FT ACT_SITE 483 483 CHARGE RELAY SYSTEM.
FT ACT_SITE 578 578 CHARGE RELAY SYSTEM.
FT DISULFID 21 104 BY SIMILARITY.
FT DISULFID 47 77 BY SIMILARITY.
FT DISULFID 51 57 BY SIMILARITY.
FT DISULFID 111 194 BY SIMILARITY.
FT DISULFID 137 166 BY SIMILARITY.
FT DISULFID 141 147 BY SIMILARITY.
FT DISULFID 201 284 BY SIMILARITY.
FT DISULFID 227 256 BY SIMILARITY.
FT DISULFID 231 237 BY SIMILARITY.
FT DISULFID 292 375 BY SIMILARITY.
FT DISULFID 318 347 BY SIMILARITY.
FT DISULFID 322 328 BY SIMILARITY.
FT DISULFID 340 345 BY SIMILARITY.
FT DISULFID 383 503 BY SIMILARITY.
FT DISULFID 419 435 BY SIMILARITY.
FT DISULFID 517 584 BY SIMILARITY.
FT DISULFID 548 563 BY SIMILARITY.
FT DISULFID 574 602 BY SIMILARITY.
SQ SEQUENCE 638 AA; 71273 MW; 454BBB27B8CABF88 CRC64;

Query Match 35.1%; Score 541; DB 1; Length 638;
Best Local Similarity 38.2%; Pred. No. 3.3e-41;
Matches 109; Conservative 54; Mismatches 112; Indels 10; Gaps 5;

1 MKLNTSAGNDVLYKLYHSDACSKAVVSLRCIACGVNLNSRROSHRIGVSGSLPAPMW 60
349 LRLSTDSGSPRTTYTAAGSSSGSLALCKVYSSDCTTKN--ARIYGTSSSLGEMW 404
61 QVSLHVNQV--HVCGGSIITPEWTVAHAQCEKPLANNMHTAFAGILROSEFMFYAGY 117
405 QVSLQVXVLSQNMHCGSSITIGROWLTAHCFD-GIPYDVWRIYGGILNLSEITWTF 463
118 -QVEVLIHPNDSTKNNDIAMQLQKPLTENDLVKPCVCLNPGMMLQPEQLCWISGHW 176
464 SSIKRLIHQKXKMSBGSYDALIKLQTEPLANTERKPCLEPSKADTNTIYTNQWTVGW 523
177 ATEEGKTSVLANAKVLLIETORQNSRYVNLITPAMICAGLOGVNDVSCQDSGGX 236
524 YTKRGFTONILQKATIFLVPEBECOKY-RDYVITKQMICAGYKGGIDACKGDSGGFL 582

```

```

QY 237 VTSKNTWMLIGDTSWGSCKAKAPRGVYGNMVTPTDITRQMA 281
DB 583 VCKSHRGMLVGLTISWBGCAKRPQGVYTKVAETIWLKTIQS 627

RESULT 10
ENTR_BOVIN STANDARD; PRT; 1035 AA.
ID ENTR_BOVIN
AC P98072;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Enteropneustidase precursor (EC 3.4.21.9) (Enterokinase).
GN PRS7 OR ENTR.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_Taxid=9913;
[1]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
TISUR=Duodenum;
RX MEDLINE=94329561; PubMed=8052624;
RA Kitamoto Y., Yuan X., Wu Q., McCourt D.W., Sadler J.E.;
RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
[2]
SEQUENCE OF 801-1035 FROM N.A., AND PARTIAL SEQUENCE.
RA Lavallie E.R., Rehmeulula A., Racie L.A., DiBlasio E.A.,
RA Perez C., Grant K.L., Light A., McCoy J.M.;
RT "Cloning and functional expression of a cDNA encoding the catalytic
subunit of bovine enterokinase."
J. Biol. Chem. 268:23311-23317(1993).
[3]
SEQUENCE OF 801-827.
RP TISSUR=Intestine;
RX MEDLINE=92189715; PubMed=1799406;
RA Light A., Janska H.;
RT "The amino-terminal sequence of the catalytic subunit of bovine
enterokinase."
J. Protein Chem. 10:475-480(1991).
-1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
PROCARBOXYPEPTIDASES, AND PROELASTASES.
-1- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-Ile-7 bond in
trypsinogen.
-1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
-1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
-1- ALTERNATIVE PRODUCTS: Type II membrane protein (Probable).
Event=Alternative splicing; Named isoforms=2;
Name=Long;
Name=Short;
CC IsoId=P98072-1; Sequence=Displayed;
CC IsoId=P98072-2; Sequence=VSP 005386;
CC -1- TISSUR SPECIFICITY: INTESTINAL BRUSH BORDER.
CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CLEAVED BY A TRYPSIN-LIKE PROTEASE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: Contains 2 LDL-receptor class A domains.
CC -1- SIMILARITY: Contains 2 CUB domains.
CC -1- SIMILARITY: Contains 1 SEA domain.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC -1- SIMILARITY: Contains 1 MAM domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

```

| Query Match | Best Local Similarity | Score | DB 1 | Length | 1035 |
|--|--|--|----------|--------|------|
| Matches | 111, Conservative | 53, Idels | 15, Gaps | 6, | |
| 1 MKLNTSANGNDVIYKKLTHSDCSKAVVSLRC--IAGCVNLNSRRQS-RIVGESALPGA 57 | 755 VMLNTAPNGSLI---LTPSQCLSDSLILQCNYSCKQLVIBVSPKIVYSGSDSREGA 811 | 34.8%, 536.5, DB 1, Length 1035; Pred. No. 1.5e-40; Mismatches 104; Indels 15; Gaps 6; | | | |
| 58 WPMQVSLFVQNVVHVGSGSIITPEVITVAHCEVKEPLNNPMHTAFAGILRQSFMYGAGY 117 | 812 WPMVVALYFDQDQVCGASVSRDLVBAHKEVYGRNPEPSKRYVGLHMASNL---TSP 868 | | | | |
| 118 QVE----KVISHPNVDSKTKNDIALMKLQKPLTFNDLVKPYCLPMPGMVLOPEOLCWI 172 | 869 QIEFRLIDQIYINHPYKRRKNDIAMHLEMKVNTYDIQPLCLPBNQVFPFGRICSI 928 | | | | |
| 173 SGMGATEBKGTSTVLAQAVLLLEIQRCKNSRYVDLLIPAMLCAGFLQGNVDSGCGDS 232 | 929 AGMVALIYQGSTADVLADADVPLLSNCKCOQOMERYN-IYENVCAGYEAGVDSGCGDS 987 | | | | |
| 233 GGLVTSKKNIMWMLIGDTSWSSGCAKRYRPGVGNVWVFTDWI 275 | 988 GGLMCGENRMLAGVTSFGYQCLAPNRFGVAVARVPEFTWI 1030 | | | | |
| <p>RESULT 11</p> <p>ENTK_HUMAN</p> <p>ID_ENTK_HUMAN STANDARD; PRT; 1019 AA.</p> <p>NCBI_98073;</p> <p>01-FEB-1996 (Rel. 33, Created)</p> <p>01-FEB-1996 (Rel. 33, Last sequence update)</p> <p>15-SEP-2003 (Rel. 42, Last annotation update)</p> <p>Enterocopepidae precursor (EC 3.4.21.9) (Enterokinase).</p> <p>PRE57 OR ENTK.</p> <p>Homo sapiens (Human).</p> <p>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.</p> <p>NCBI_taxonomy:9606;</p> <p>(1)</p> <p>SEQUENCE FROM N.A.</p> <p>TISSUE=Duodenum;</p> <p>MEDLINE=85234679; PubMed=7718557;</p> <p>Kitamoto Y., Velle R.A., Donis-Reller H., Sadler J.B.;</p> <p>"cDNA sequence and chromosomal localization of human enterokinase, the proteolytic activator of trypsinogen.",</p> <p>Biochemistry 34:4562-4568(1995).</p> <p>(2)</p> <p>SEQUENCE FROM N.A. AND DISTANCE</p> | | | | | |

RX MEDLINE=21606074; PubMed=11719902;
RA Holzinger A., Maier E.M., Buck C., Mayerhofer P.U., Kappler M.,
RA Haworth J.C., Moroz S.P., Hadorn H.-B., Sadler J.E., Roscher A.A.,
RT "Mutations in the proenteropeptidase gene are the molecular cause of
RT congenital enteropeptidase deficiency.";
RL Am. J. Hum. Genet. 70:20-25(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichenwald K., Kump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Akakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuhashi S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordle K., Horikawa K., Brandt P.,
RA Scharte M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecher H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riesemann L., Dagand E.,
RA Weinmayer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehnach H., Reinhardt R., Yaspo M.-L.,
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN [4]
RP SEQUENCE OF 749-1019 FROM N.A.
RX TISSUE=Duodenum;
RX MEDLINE=94329561; PubMed=8052624;
RA Kitamoto Y., Yuan X., Wu Q., McCourt D.W., Sadler J.E.,
RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
RT protease composed of a distinctive assortment of domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
CC A). IF CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
CC PROCARBOXYPEPTIDASES, AND PROELASTASES.
CC -1- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-|-Ile-7 bond in
CC trypsinogen.
CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -1- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPSIN-LIKE PROTEASE.
CC -1- DISEASE: Defects in PRSS7 are a cause of enterokinase deficiency
CC [MIM:226200]; a life-threatening intestinal malabsorption disorder
CC characterized by diarrhea and failure to thrive.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: Contains 2 LDL-receptor class A domains.
CC -1- SIMILARITY: Contains 2 CUB domains.
CC -1- SIMILARITY: Contains 1 SEA domain.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC -1- SIMILARITY: Contains 1 MAM domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; U09860; AAC50138.1; -
CC EMBL; Y19124; CAB65555.1; JOINED.
CC EMBL; Y19125; CAB65555.1; JOINED.
CC EMBL; Y19126; CAB65555.1; JOINED.
CC EMBL; Y19127; CAB65555.1; JOINED.
CC EMBL; Y19128; CAB65555.1; JOINED.
CC EMBL; Y19129; CAB65555.1; JOINED.
CC EMBL; Y19130; CAB65555.1; JOINED.
CC EMBL; Y19131; CAB65555.1; JOINED.
CC EMBL; Y19132; CAB65555.1; JOINED.
CC EMBL; Y19133; CAB65555.1; JOINED.

DR EMBL; Y19134; CAB65555.1; JOINED.
DR EMBL; Y19135; CAB65555.1; JOINED.
DR EMBL; Y19136; CAB65555.1; JOINED.
DR EMBL; Y19137; CAB65555.1; JOINED.
DR EMBL; Y19138; CAB65555.1; JOINED.
DR EMBL; Y19139; CAB65555.1; JOINED.
DR EMBL; Y19140; CAB65555.1; JOINED.
DR EMBL; Y19141; CAB65555.1; JOINED.
DR EMBL; Y19142; CAB65555.1; JOINED.
DR EMBL; Y19143; CAB65555.1; JOINED.
DR EMBL; Y19144; CAB65555.1; JOINED.
DR EMBL; Y19145; CAB65555.1; JOINED.
DR EMBL; Y19146; CAB65555.1; JOINED.
DR EMBL; Y19147; CAB65555.1; JOINED.
DR EMBL; Y19148; CAB65555.1; JOINED.
DR EMBL; Y19149; CAB65555.1; JOINED.
DR EMBL; Y19150; CAB65555.1; JOINED.
DR EMBL; Y19151; CAB65555.1; JOINED.
DR EMBL; Y19152; CAB65555.1; JOINED.
DR EMBL; Y19153; CAB65555.1; JOINED.
DR EMBL; Y19154; CAB65555.1; JOINED.
DR EMBL; Y19155; CAB65555.1; JOINED.
DR EMBL; Y19156; CAB65555.1; JOINED.
DR EMBL; Y19157; CAB65555.1; JOINED.
DR EMBL; Y19158; CAB65555.1; JOINED.
DR EMBL; Y19159; CAB65555.1; JOINED.
DR EMBL; Y19160; CAB65555.1; JOINED.
DR EMBL; Y19161; CAB65555.1; JOINED.
DR EMBL; Y19162; CAB65555.1; JOINED.
DR EMBL; Y19163; CAB65555.1; JOINED.
DR EMBL; Y19164; CAB65555.1; JOINED.
DR EMBL; Y19165; CAB65555.1; JOINED.
DR EMBL; Y19166; CAB65555.1; JOINED.
DR EMBL; Y19167; CAB65555.1; JOINED.
DR EMBL; Y19168; CAB65555.1; JOINED.
DR EMBL; Y19169; CAB65555.1; JOINED.
DR EMBL; Y19170; CAB65555.1; JOINED.
DR EMBL; Y19171; CAB65555.1; JOINED.
DR EMBL; Y19172; CAB65555.1; JOINED.
DR EMBL; Y19173; CAB65555.1; JOINED.
DR EMBL; Y19174; CAB65555.1; JOINED.
DR EMBL; Y19175; CAB65555.1; JOINED.
DR EMBL; Y19176; CAB65555.1; JOINED.
DR EMBL; Y19177; CAB65555.1; JOINED.
DR EMBL; Y19178; CAB65555.1; JOINED.
DR EMBL; Y19179; CAB65555.1; JOINED.
DR EMBL; Y19180; CAB65555.1; JOINED.
DR EMBL; Y19181; CAB65555.1; JOINED.
DR EMBL; Y19182; CAB65555.1; JOINED.
DR EMBL; Y19183; CAB65555.1; JOINED.
DR EMBL; Y19184; CAB65555.1; JOINED.
DR EMBL; Y19185; CAB65555.1; JOINED.
DR EMBL; Y19186; CAB65555.1; JOINED.
DR EMBL; Y19187; CAB65555.1; JOINED.
DR EMBL; Y19188; CAB65555.1; JOINED.
DR EMBL; Y19189; CAB65555.1; JOINED.
DR EMBL; Y19190; CAB65555.1; JOINED.
DR EMBL; Y19191; CAB65555.1; JOINED.
DR EMBL; Y19192; CAB65555.1; JOINED.
DR EMBL; Y19193; CAB65555.1; JOINED.
DR EMBL; Y19194; CAB65555.1; JOINED.
DR EMBL; Y19195; CAB65555.1; JOINED.
DR EMBL; Y19196; CAB65555.1; JOINED.
DR EMBL; Y19197; CAB65555.1; JOINED.
DR EMBL; Y19198; CAB65555.1; JOINED.
DR EMBL; Y19199; CAB65555.1; JOINED.
DR EMBL; Y19200; CAB65555.1; JOINED.
DR EMBL; Y19201; CAB65555.1; JOINED.
DR EMBL; Y19202; CAB65555.1; JOINED.
DR EMBL; Y19203; CAB65555.1; JOINED.
DR EMBL; Y19204; CAB65555.1; JOINED.
DR EMBL; Y19205; CAB65555.1; JOINED.
DR EMBL; Y19206; CAB65555.1; JOINED.
DR EMBL; Y19207; CAB65555.1; JOINED.
DR EMBL; Y19208; CAB65555.1; JOINED.
DR EMBL; Y19209; CAB65555.1; JOINED.
DR EMBL; Y19210; CAB65555.1; JOINED.
DR EMBL; Y19211; CAB65555.1; JOINED.
DR EMBL; Y19212; CAB65555.1; JOINED.
DR EMBL; Y19213; CAB65555.1; JOINED.
DR EMBL; Y19214; CAB65555.1; JOINED.
DR EMBL; Y19215; CAB65555.1; JOINED.
DR EMBL; Y19216; CAB65555.1; JOINED.
DR EMBL; Y19217; CAB65555.1; JOINED.
DR EMBL; Y19218; CAB65555.1; JOINED.
DR EMBL; Y19219; CAB65555.1; JOINED.
DR EMBL; Y19220; CAB65555.1; JOINED.
DR EMBL; Y19221; CAB65555.1; JOINED.
DR EMBL; Y19222; CAB65555.1; JOINED.
DR EMBL; Y19223; CAB65555.1; JOINED.
DR EMBL; Y19224; CAB65555.1; JOINED.
DR EMBL; Y19225; CAB65555.1; JOINED.
DR EMBL; Y19226; CAB65555.1; JOINED.
DR EMBL; Y19227; CAB65555.1; JOINED.
DR EMBL; Y19228; CAB65555.1; JOINED.
DR EMBL; Y19229; CAB65555.1; JOINED.
DR EMBL; Y19230; CAB65555.1; JOINED.
DR EMBL; Y19231; CAB65555.1; JOINED.
DR EMBL; Y19232; CAB65555.1; JOINED.
DR EMBL; Y19233; CAB65555.1; JOINED.
DR EMBL; Y19234; CAB65555.1; JOINED.
DR EMBL; Y19235; CAB65555.1; JOINED.
DR EMBL; Y19236; CAB65555.1; JOINED.
DR EMBL; Y19237; CAB65555.1; JOINED.
DR EMBL; Y19238; CAB65555.1; JOINED.
DR EMBL; Y19239; CAB65555.1; JOINED.
DR EMBL; Y19240; CAB65555.1; JOINED.
DR EMBL; Y19241; CAB65555.1; JOINED.
DR EMBL; Y19242; CAB65555.1; JOINED.
DR EMBL; Y19243; CAB65555.1; JOINED.
DR EMBL; Y19244; CAB65555.1; JOINED.
DR EMBL; Y19245; CAB65555.1; JOINED.
DR EMBL; Y19246; CAB65555.1; JOINED.
DR EMBL; Y19247; CAB65555.1; JOINED.
DR EMBL; Y19248; CAB65555.1; JOINED.
DR EMBL; Y19249; CAB65555.1; JOINED.
DR EMBL; Y19250; CAB65555.1; JOINED.
DR EMBL; Y19251; CAB65555.1; JOINED.
DR EMBL; Y19252; CAB65555.1; JOINED.
DR EMBL; Y19253; CAB65555.1; JOINED.
DR EMBL; Y19254; CAB65555.1; JOINED.
DR EMBL; Y19255; CAB65555.1; JOINED.
DR EMBL; Y19256; CAB65555.1; JOINED.
DR EMBL; Y19257; CAB65555.1; JOINED.
DR EMBL; Y19258; CAB65555.1; JOINED.
DR EMBL; Y19259; CAB65555.1; JOINED.
DR EMBL; Y19260; CAB65555.1; JOINED.
DR EMBL; Y19261; CAB65555.1; JOINED.
DR EMBL; Y19262; CAB65555.1; JOINED.
DR EMBL; Y19263; CAB65555.1; JOINED.
DR EMBL; Y19264; CAB65555.1; JOINED.
DR EMBL; Y19265; CAB65555.1; JOINED.
DR EMBL; Y19266; CAB65555.1; JOINED.
DR EMBL; Y19267; CAB65555.1; JOINED.
DR EMBL; Y19268; CAB65555.1; JOINED.
DR EMBL; Y19269; CAB65555.1; JOINED.
DR EMBL; Y19270; CAB65555.1; JOINED.
DR EMBL; Y19271; CAB65555.1; JOINED.
DR EMBL; Y19272; CAB65555.1; JOINED.
DR EMBL; Y19273; CAB65555.1; JOINED.
DR EMBL; Y19274; CAB65555.1; JOINED.
DR EMBL; Y19275; CAB65555.1; JOINED.
DR EMBL; Y19276; CAB65555.1; JOINED.
DR EMBL; Y19277; CAB65555.1; JOINED.
DR EMBL; Y19278; CAB65555.1; JOINED.
DR EMBL; Y19279; CAB65555.1; JOINED.
DR EMBL; Y19280; CAB65555.1; JOINED.
DR EMBL; Y19281; CAB65555.1; JOINED.
DR EMBL; Y19282; CAB65555.1; JOINED.
DR EMBL; Y19283; CAB65555.1; JOINED.
DR EMBL; Y19284; CAB65555.1; JOINED.
DR EMBL; Y19285; CAB65555.1; JOINED.
DR EMBL; Y19286; CAB65555.1; JOINED.
DR EMBL; Y19287; CAB65555.1; JOINED.
DR EMBL; Y19288; CAB65555.1; JOINED.
DR EMBL; Y19289; CAB65555.1; JOINED.
DR EMBL; Y19290; CAB65555.1; JOINED.
DR EMBL; Y19291; CAB65555.1; JOINED.
DR EMBL; Y19292; CAB65555.1; JOINED.
DR EMBL; Y19293; CAB65555.1; JOINED.
DR EMBL; Y19294; CAB65555.1; JOINED.
DR EMBL; Y19295; CAB65555.1; JOINED.
DR EMBL; Y19296; CAB65555.1; JOINED.
DR EMBL; Y19297; CAB65555.1; JOINED.
DR EMBL; Y19298; CAB65555.1; JOINED.
DR EMBL; Y19299; CAB65555.1; JOINED.
DR EMBL; Y19300; CAB65555.1; JOINED.
DR EMBL; Y19301; CAB65555.1; JOINED.
DR EMBL; Y19302; CAB65555.1; JOINED.
DR EMBL; Y19303; CAB65555.1; JOINED.
DR EMBL; Y19304; CAB65555.1; JOINED.
DR EMBL; Y19305; CAB65555.1; JOINED.
DR EMBL; Y19306; CAB65555.1; JOINED.
DR EMBL; Y19307; CAB65555.1; JOINED.
DR EMBL; Y19308; CAB65555.1; JOINED.
DR EMBL; Y19309; CAB65555.1; JOINED.
DR EMBL; Y19310; CAB65555.1; JOINED.
DR EMBL; Y19311; CAB65555.1; JOINED.
DR EMBL; Y19312; CAB65555.1; JOINED.
DR EMBL; Y19313; CAB65555.1; JOINED.
DR EMBL; Y19314; CAB65555.1; JOINED.
DR EMBL; Y19315; CAB65555.1; JOINED.
DR EMBL; Y19316; CAB65555.1; JOINED.
DR EMBL; Y19317; CAB65555.1; JOINED.
DR EMBL; Y19318; CAB65555.1; JOINED.
DR EMBL; Y19319; CAB65555.1; JOINED.
DR EMBL; Y19320; CAB65555.1; JOINED.
DR EMBL; Y19321; CAB65555.1; JOINED.
DR EMBL; Y19322; CAB65555.1; JOINED.
DR EMBL; Y19323; CAB65555.1; JOINED.
DR EMBL; Y19324; CAB65555.1; JOINED.
DR EMBL; Y19325; CAB65555.1; JOINED.
DR EMBL; Y19326; CAB65555.1; JOINED.
DR EMBL; Y19327; CAB65555.1; JOINED.
DR EMBL; Y19328; CAB65555.1; JOINED.
DR EMBL; Y19329; CAB65555.1; JOINED.
DR EMBL; Y19330; CAB65555.1; JOINED.
DR EMBL; Y19331; CAB65555.1; JOINED.
DR EMBL; Y19332; CAB65555.1; JOINED.
DR EMBL; Y19333; CAB65555.1; JOINED.
DR EMBL; Y19334; CAB65555.1; JOINED.
DR EMBL; Y19335; CAB65555.1; JOINED.
DR EMBL; Y19336; CAB65555.1; JOINED.
DR EMBL; Y19337; CAB65555.1; JOINED.
DR EMBL; Y19338; CAB65555.1; JOINED.
DR EMBL; Y19339; CAB65555.1; JOINED.
DR EMBL; Y19340; CAB65555.1; JOINED.
DR EMBL; Y19341; CAB65555.1; JOINED.
DR EMBL; Y19342; CAB65555.1; JOINED.
DR EMBL; Y19343; CAB65555.1; JOINED.
DR EMBL; Y19344; CAB65555.1; JOINED.
DR EMBL; Y19345; CAB65555.1; JOINED.
DR EMBL; Y19346; CAB65555.1; JOINED.
DR EMBL; Y19347; CAB65555.1; JOINED.
DR EMBL; Y19348; CAB65555.1; JOINED.
DR EMBL; Y19349; CAB65555.1; JOINED.
DR EMBL; Y19350; CAB65555.1; JOINED.
DR EMBL; Y19351; CAB65555.1; JOINED.
DR EMBL; Y19352; CAB65555.1; JOINED.
DR EMBL; Y19353; CAB65555.1; JOINED.
DR EMBL; Y19354; CAB65555.1; JOINED.
DR EMBL; Y19355; CAB65555.1; JOINED.
DR EMBL; Y19356; CAB65555.1; JOINED.
DR EMBL; Y19357; CAB65555.1; JOINED.
DR EMBL; Y19358; CAB65555.1; JOINED.
DR EMBL; Y19359; CAB65555.1; JOINED.
DR EMBL; Y19360; CAB65555.1; JOINED.
DR EMBL; Y19361; CAB65555.1; JOINED.
DR EMBL; Y19362; CAB65555.1; JOINED.
DR EMBL; Y19363; CAB65555.1; JOINED.
DR EMBL; Y19364; CAB65555.1; JOINED.
DR EMBL; Y19365; CAB65555.1; JOINED.
DR EMBL; Y19366; CAB65555.1; JOINED.
DR EMBL; Y19367; CAB65555.1; JOINED.
DR EMBL; Y19368; CAB65555.1; JOINED.
DR EMBL; Y19369; CAB65555.1; JOINED.
DR EMBL; Y19370; CAB65555.1; JOINED.
DR EMBL; Y19371; CAB65555.1; JOINED.
DR EMBL; Y19372; CAB65555.1; JOINED.
DR EMBL; Y19373; CAB65555.1; JOINED.
DR EMBL; Y19374; CAB65555.1; JOINED.
DR EMBL; Y19375; CAB65555.1; JOINED.
DR EMBL; Y19376; CAB65555.1; JOINED.
DR EMBL; Y19377; CAB65555.1; JOINED.
DR EMBL; Y19378; CAB65555.1; JOINED.
DR EMBL; Y19379; CAB65555.1; JOINED.
DR EMBL; Y19380; CAB65555.1; JOINED.
DR EMBL; Y19381; CAB65555.1; JOINED.
DR EMBL; Y19382; CAB65555.1; JOINED.
DR EMBL; Y19383; CAB65555.1; JOINED.
DR EMBL; Y19384; CAB65555.1; JOINED.
DR EMBL; Y19385; CAB65555.1; JOINED.
DR EMBL; Y19386; CAB65555.1; JOINED.
DR EMBL; Y19387; CAB65555.1; JOINED.
DR EMBL; Y19388; CAB65555.1; JOINED.
DR EMBL; Y19389; CAB65555.1; JOINED.
DR EMBL; Y19390; CAB65555.1; JOINED.
DR EMBL; Y19391; CAB65555.1; JOINED.
DR EMBL; Y19392; CAB65555.1; JOINED.
DR EMBL; Y19393; CAB65555.1; JOINED.
DR EMBL; Y19394; CAB65555.1; JOINED.
DR EMBL; Y19395; CAB65555.1; JOINED.
DR EMBL; Y19396; CAB65555.1; JOINED.
DR EMBL; Y19397; CAB65555.1; JOINED.
DR EMBL; Y19398; CAB65555.1; JOINED.
DR EMBL; Y19399; CAB65555.1; JOINED.
DR EMBL; Y19400; CAB65555.1; JOINED.
DR EMBL; Y19401; CAB65555.1; JOINED.
DR EMBL; Y19402; CAB65555.1; JOINED.
DR EMBL; Y19403; CAB65555.1; JOINED.
DR EMBL; Y19404; CAB65555.1; JOINED.
DR EMBL; Y19405; CAB65555.1; JOINED.
DR EMBL; Y19406; CAB65555.1; JOINED.
DR EMBL; Y19407; CAB65555.1; JOINED.
DR EMBL; Y19408; CAB65555.1; JOINED.
DR EMBL; Y19409; CAB65555.1; JOINED.
DR EMBL; Y19410; CAB65555.1; JOINED.
DR EMBL; Y19411; CAB65555.1; JOINED.
DR EMBL; Y19412; CAB65555.1; JOINED.
DR EMBL; Y19413; CAB65555.1; JOINED.
DR EMBL; Y19414; CAB65555.1; JOINED.
DR EMBL; Y19415; CAB65555.1; JOINED.
DR EMBL; Y19416; CAB65555.1; JOINED.
DR EMBL; Y19417; CAB65555.1; JOINED.
DR EMBL; Y19418; CAB65555.1; JOINED.
DR EMBL; Y19419; CAB65555.1; JOINED.
DR EMBL; Y19420; CAB65555.1; JOINED.
DR EMBL; Y19421; CAB65555.1; JOINED.
DR EMBL; Y19422; CAB65555.1; JOINED.
DR EMBL; Y19423; CAB65555.1; JOINED.
DR EMBL; Y19424; CAB65555.1; JOINED.
DR EMBL; Y19425; CAB65555.1; JOINED.
DR EMBL; Y19426; CAB65555.1; JOINED.
DR EMBL; Y19427; CAB65555.1; JOINED.
DR EMBL; Y19428; CAB65555.1; JOINED.
DR EMBL; Y19429; CAB65555.1; JOINED.
DR EMBL; Y19430; CAB65555.1; JOINED.
DR EMBL; Y19431; CAB65555.1; JOINED.
DR EMBL; Y19432; CAB65555.1; JOINED.
DR EMBL; Y19433; CAB65555.1; JOINED.
DR EMBL; Y19434; CAB65555.1; JOINED.
DR EMBL; Y19435; CAB65555.1; JOINED.
DR EMBL; Y19436; CAB65555.1; JOINED.
DR EMBL; Y19437; CAB65555.1; JOINED.
DR EMBL; Y19438; CAB65555.1; JOINED.
DR EMBL; Y19439; CAB65555.1; JOINED.
DR EMBL; Y19440; CAB65555.1; JOINED.
DR EMBL; Y19441; CAB65555.1; JOINED.
DR EMBL; Y19442; CAB65555.1; JOINED.
DR EMBL; Y19443; CAB65555.1; JOINED.
DR EMBL; Y19444; CAB65555.1; JOINED.
DR EMBL; Y19445; CAB65555.1; JOINED.
DR EMBL; Y19446; CAB65555.1; JOINED.
DR EMBL; Y19447; CAB65555.1; JOINED.
DR EMBL; Y19448; CAB65555.1; JOINED.
DR EMBL; Y19449; CAB65555.1; JOINED.
DR EMBL; Y19450; CAB65555.1; JOINED.
DR EMBL; Y19451; CAB65555.1; JOINED.
DR EMBL; Y19452; CAB65555.1; JOINED.
DR EMBL; Y19453; CAB65555.1; JOINED.
DR EMBL; Y19454; CAB65555.1; JOINED.
DR EMBL; Y19455; CAB65555.1; JOINED.
DR EMBL; Y19456; CAB65555.1; JOINED.
DR EMBL; Y19457; CAB65555.1; JOINED.
DR EMBL; Y19458; CAB65555.1; JOINED.
DR EMBL; Y19459; CAB65555.1; JOINED.
DR EMBL; Y19460; CAB65555.1; JOINED.
DR EMBL; Y19461; CAB65555.1; JOINED.
DR EMBL; Y19462; CAB65555.1; JOINED.
DR EMBL; Y19463; CAB65555.1; JOINED.
DR EMBL; Y19464; CAB65555.1; JOINED.
DR EMBL; Y19465; CAB65555.1; JOINED.
DR EMBL; Y19466; CAB65555.1; JOINED.
DR EMBL; Y19467; CAB65555.1; JOINED.
DR EMBL; Y19468; CAB65555.1; JOINED.
DR EMBL; Y19469; CAB65555.1; JOINED.
DR EMBL; Y19470; CAB65555.1; JOINED.
DR EMBL; Y19471; CAB65555.1; JOINED.
DR EMBL; Y19472; CAB65555.1; JOINED.
DR EMBL; Y19473; CAB65555.1; JOINED.
DR EMBL; Y19474; CAB65555.1; JOINED.
DR EMBL; Y19475; CAB65555.1; JOINED.
DR EMBL; Y19476; CAB65555.1; JOINED.
DR EMBL; Y19477; CAB65555.1; JOINED.
DR EMBL; Y19478; CAB65555.1; JOINED.
DR EMBL; Y19479; CAB65555.1; JOINED.
DR EMBL; Y19480; CAB65555.1; JOINED.
DR EMBL; Y19481; CAB65555.1; JOINED.
DR EMBL; Y19482; CAB65555.1; JOINED.
DR EMBL; Y19483; CAB65555.1; JOINED.
DR EMBL; Y19484; CAB65555.1; JOINED.
DR EMBL; Y19485; CAB65555.1; JOINED.
DR EMBL; Y19486; CAB65555.1; JOINED.
DR EMBL; Y19487; CAB65555.1; JOINED.
DR EMBL; Y19488; CAB65555.1; JOINED.
DR EMBL; Y19489; CAB65555.1; JOINED.
DR EMBL; Y19490; CAB65555.1; JOINED.
DR EMBL; Y19491; CAB65555.1; JOINED.
DR EMBL; Y19492; CAB65555.1; JOINED.
DR EMBL; Y19493; CAB65555.1; JOINED.
DR EMBL; Y19494; CAB65555.1; JOINED.
DR EMBL; Y19495; CAB65555.1; JOINED.
DR EMBL; Y19496; CAB65555.1; JOINED.
DR EMBL; Y19497; CAB65555.1; JOINED.
DR EMBL; Y19498; CAB65555.1; JOINED.
DR EMBL; Y19499; CAB65555.1; JOINED.
DR EMBL; Y19500; CAB65555.1; JOINED.
DR EMBL; Y19501; CAB65555.1; JOINED.
DR EMBL; Y19502; CAB65555.1; JOINED.
DR EMBL; Y19503; CAB65555.1; JOINED.
DR EMBL; Y19504; CAB65555.1; JOINED.
DR EMBL; Y19505; CAB65555.1; JOINED.
DR EMBL; Y19506; CAB65555.1; JOINED.
DR EMBL; Y19507; CAB65555.1; JOINED.
DR EMBL; Y19508; CAB65555.1; JOINED.
DR EMBL; Y19509; CAB65555.1; JOINED.
DR EMBL; Y19510; CAB65555.1; JOINED.
DR EMBL; Y19511; CAB65555.1; JOINED.
DR EMBL; Y19512; CAB65555.1; JOINED.
DR EMBL; Y19513; CAB65555.1; JOINED.
DR EMBL; Y19514; CAB65555.1; JOINED.
DR EMBL; Y19515; CAB65555.1; JOINED.
DR EMBL; Y19516; CAB65555.1; JOINED.
DR EMBL; Y19517; CAB65555.1; JOINED.
DR EMBL; Y19518; CAB65555.1; JOINED.
DR EMBL; Y19519; CAB65555.1; JOINED.
DR EMBL; Y19520; CAB65555.1; JOINED.
DR EMBL; Y19521; CAB65555.1; JOINED.
DR EMBL; Y19522; CAB65555.1; JOINED.
DR EMBL; Y19523; CAB65555.1;

```

FT DISULFID 191 210 BY SIMILARITY.
FT DISULFID 204 221 BY SIMILARITY.
FT DISULFID 643 655 BY SIMILARITY.
FT DISULFID 650 668 BY SIMILARITY.
FT DISULFID 662 677 BY SIMILARITY.
FT DISULFID 772 896 INTERCHAIN (BY SIMILARITY).
FT DISULFID 810 826 BY SIMILARITY.
FT DISULFID 910 977 BY SIMILARITY.
FT DISULFID 941 956 BY SIMILARITY.
FT CARBOHYD 967 995 BY SIMILARITY.
FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 328 328 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 335 335 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 388 388 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 440 440 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 503 503 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 630 630 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 682 682 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 706 706 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 848 848 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 887 887 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 909 909 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 949 949 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 134 134 Q -> B (IN REF. 3).
FT CONFLICT 134 134 S -> P (IN REF. 3).
FT CONFLICT 732 732 SOOCLDSDIRLCQCNHKS -> RRNAKEIDALSPITLLA
FT CONFLICT 754 771 (IN REF. 3).
SQ SEQUENCE 1019 AA; 112923 MW; 66AA245F6DA563 CRC64;

Query Match 34.3%; Score 528.5; DB 1; length 1019;
Best Local Similarity 38.6%; Pred. No. 7.5e-40;
Matches 108; Conservative 56; Mismatches 107; Indels 9; Gaps 5;

OY 1 MKLNTSAGNVDDYKLYHSDDACSKAVVSLRC--IACGVNINSSR-QSRIVGESLPLPA 57
DB 739 VKLNTAP---DHLILITPSOCLDSDIRLCQCNHKSCKKLAQDITPPIVGSNNKEGA 795
OY 58 WPMOVSLHVNQNVHVGSGSITTEPWITAAHCEKPLANPMHTAFAGILROSEFMF--YGA 115
DB 796 WPMVVGGLYYGGRLLCGASIVSSDMVSAHCVYGRULEPSKXTALIGLHKSLTSPQVY 855
OY 116 GYQVEKVISHPNYSKTKNDIALMKLQKPLTFNDLVKVCCLPNPGMLOPQOLCWIISGM 175
DB 856 PRLIDEIYINPHYNNRRKNDIAMMHLERKVNVTYDIPICLPEBNQVPPPGNCSIGAM 915
OY 176 GATEBKGTSEVLNAKYLITFORCNSRYVDNLITPMICAGPIQGVNDSCQDSGCGX 235
DB 916 GTVVYGGTATNLLQADVPILLSNERCQQQMPBYN-ITEMNICGVEEGIDSCQDSGSGP 974
OY 236 LVTSKNNTIMWLIGTSMGSGCAKAYRPGVYGNVVTDWI 275
DB 975 LMCQENNRHPLAGVTSFGYKCALPMRPGVYARVASFTEWI 1014

RESULT 12
KAL_MOUSE STANDARD; PRT; 638 AA.
AC P26362
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Plasma kallikrein precursor (BC 3.4.21.34) (Plasma prekallikrein)
DE (Kininogenin) (Fletcher factor).
DE KLK3 OR KLK3 OR PK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

```

```

RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-BALB/C; Tissue=Liver;
RX MEDLINE=91090844; PubMed=2264928;
RA Seidan N.G., Sawyer N., Hamelin J., Mion P., Beaudien G.,
RA Brachpapa L., Rochemont J., Mblay M., Chretien M.,
RT "Mouse plasma kallikrein: cDNA structure, enzyme characterization,
RT and comparison of protein and mRNA levels among species.";
RL DNA Cell Biol. 9:737-748(1990).

CC -1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
CC -1- CATALYTIC ACTIVITY: Cleaves selectively Arg-|-Xaa and Lys-|-Xaa
CC bonds, including Lys-|-Arg and Arg-|-Ser bonds in (human)
CC kininogen to release bradykinin.
CC -1- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XIIa, WHICH CLEAVES
CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. PLASMA KALLIKREIN
CC SUBFAMILY.
CC -1- SIMILARITY: Contains 4 apple domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.1eb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M58588; AAA63393.1; -.
DR PIR; A36557; KQMSPL.
DR HSSP; P00750; IRTF.
DR MEROPS; S01.212; -.
DR MGD; MGI:102849; Klkb1.
DR InterPro; IPR000177; Apple.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR003014; PAN.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00024; PAN; 4.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00005; APPLEDOMAIN.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00223; APPLE; 4.
DR SMART; SM00200; TYP_SPC; 1.
DR PROSITE; PS00495; APPLE; 4.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
KW Fibinolysis; Blood coagulation; Inflammatory response; Liver;
KW Repeat.
FT SIGNAL 1 19
FT CHAIN 20 390
FT CHAIN 391 638
FT DOMAIN 20 105
FT DOMAIN 110 195
FT DOMAIN 200 285
FT DOMAIN 291 376
FT DOMAIN 389 621
FT CARBOHYD 127 127
FT CARBOHYD 308 308
FT CARBOHYD 396 396
FT CARBOHYD 453 453
FT CARBOHYD 454 454
FT ACT_SITE 494 494
FT ACT_SITE 483 483
FT ACT_SITE 483 483
FT ACT_SITE 578 578
FT DISULFID 21 104
FT DISULFID 47 77

PLASMA KALLIKREIN HEAVY CHAIN.
PLASMA KALLIKREIN LIGHT CHAIN.
APPLE-1.
APPLE 2.
APPLE 3.
APPLE 4.
SERINE PROTEASE.
N-LINKED (GLCNAC. . .) (PROBABLE).
N-LINKED (GLCNAC. . .) (PROBABLE).
N-LINKED (GLCNAC. . .) (PROBABLE).
O-LINKED (PROBABLE. . .) (PROBABLE).
N-LINKED (GLCNAC. . .) (PROBABLE).
CHARGE RELAY SYSTEM.
CHARGE RELAY SYSTEM.
CHARGE RELAY SYSTEM.
BY SIMILARITY.
BY SIMILARITY.

```

```

FT DISULFID 51 57 BY SIMILARITY.
FT DISULFID 111 194 BY SIMILARITY.
FT DISULFID 137 166 BY SIMILARITY.
FT DISULFID 141 147 BY SIMILARITY.
FT DISULFID 201 284 BY SIMILARITY.
FT DISULFID 227 256 BY SIMILARITY.
FT DISULFID 231 237 BY SIMILARITY.
FT DISULFID 292 375 BY SIMILARITY.
FT DISULFID 318 347 BY SIMILARITY.
FT DISULFID 322 328 BY SIMILARITY.
FT DISULFID 340 345 BY SIMILARITY.
FT DISULFID 383 503 BY SIMILARITY.
FT DISULFID 419 435 BY SIMILARITY.
FT DISULFID 517 584 BY SIMILARITY.
FT DISULFID 548 563 BY SIMILARITY.
FT DISULFID 574 602 BY SIMILARITY.
SQ SEQUENCE 638 AA; 71368 MW; CC27C93AB1086599 CR664;

Query Match 34.1%; Score 526.5; DB 1; Length 638;
Best Local Similarity 39.1%; Pred. No. 6,7e-40;
Matches 106; Conservative 46; Mismatches 96; Indels 23; Gaps 5;

QY 15 KLVHSDACSSKAVVSLRCIACGVNLSRSRSHVGSALPAMPPOVSIHVQNV---HV 71
Db KLVHSDACSSKAVVSLRCIACGVNLSRSRSHVGSALPAMPPOVSIHVQNV---HV 71
QY 72 CGGSIITPEWITVAHCVCEKPLNPMHTAFAGILRQSFMGAG-GYOVKVISHPNDS 130
Db CGGSIITPEWITVAHCVCEKPLNPMHTAFAGILRQSFMGAG-GYOVKVISHPNDS 130
QY 419 CGGSIITPEWITVAHCVCEKPLNPMHTAFAGILRQSFMGAG-GYOVKVISHPNDS 130
Db CGGSIITPEWITVAHCVCEKPLNPMHTAFAGILRQSFMGAG-GYOVKVISHPNDS 130
QY 131 KTRKNDITMLQKRLTNDLVKPCVLPKPGMLQPEOLCISGKATREKTSVANA 190
Db KTRKNDITMLQKRLTNDLVKPCVLPKPGMLQPEOLCISGKATREKTSVANA 190
QY 478 SEGNYDALIKLQPLNTEPKPCIPSKADTNTITNTCWVTGWGKYEQSEYNTIOLK 537
Db SEGNYDALIKLQPLNTEPKPCIPSKADTNTITNTCWVTGWGKYEQSEYNTIOLK 537
QY 191 AKVLIETORCNSRVYDNLITPAMICAGFLQGVNDSGCGGKLVTSKNNIMWILGPT 250
Db AKVLIETORCNSRVYDNLITPAMICAGFLQGVNDSGCGGKLVTSKNNIMWILGPT 250
QY 538 ATILVPEBECOKK-RDVIYNKMKICAGYKGGGADCKDGGGGLVCHGSRMQLVGT 596
Db ATILVPEBECOKK-RDVIYNKMKICAGYKGGGADCKDGGGGLVCHGSRMQLVGT 596
QY 251 SWSGGACAVAPRGVYGVNVTFTWIRQMA 281
Db SWSGGACAVAPRGVYGVNVTFTWIRQMA 281
QY 597 SWGEGCGRKDQPGVYTKVSEVMDILKTKQS 627
Db SWGEGCGRKDQPGVYTKVSEVMDILKTKQS 627

RESULT 13
HEPS_HUMAN STANDARD; PRT; 417 AA.
AC P05981;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Serine protease hepsin (EC 3.4.21.-) (Transmembrane protease, serine
1).
GN HPN OR TMRSS1.
OS Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCB1_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=86209431; PubMed=2835076;
RA Leytus S.P., Loeb K.R., Hagen F.S., Kurachi K., Davie E.W.;
RT "A novel trypsin-like serine protease (hepsin) with a putative
transmembrane domain expressed by human liver and hepatoma cells.";
RL Biochemistry 27:1067-1074(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreeas, and Spleen;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heich F.,

```

```

RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stappleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein W.J., Usdin T.B., Toshiyuki S., Carmichael P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosnak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren B.J., Lu X., Gibbs R.A.,
RA Pahay J., Helton B., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalske U., Smalins D.B.,
RA Scherch A., Schein J.B., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=91358502; PubMed=1885621;
RA Tsuiji A., Torres-Rosado A., Arai T., le Beau M.M., Lemons R.S.,
RA Chou S.H., Kurachi K.;
RT "Hepsin, a cell membrane-associated protease. Characterization,
RT tissue distribution, and gene localization.";
RL J. Biol. Chem. 266:16948-16953(1991).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=93348237; PubMed=8346233;
RA Torres-Rosado A., O'Shea K.S., Tsuiji A., Chou S.H., Kurachi K.;
RT "Hepsin, a putative cell-surface serine protease, is required for
RT mammalian cell growth.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:7181-7187(1993).
CC -1- FUNCTION: Plays an essential role in cell growth and maintenance
CC of cell morphology.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
CC -1- TISSUE SPECIFICITY: Present in most tissues, with the highest
CC level in liver.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL, M18930; AAA36013.1; -.
DR EMBL, X07732; CA330558.1; -.
DR EMBL, X07002; CA330058.1; -.
DR EMBL, BC025716; AA425716.1; -.
DR PIR, S00845; S00845.
DR HSSP, P00763; IDPO.
DR MEROPS, S01.224; -.
DR GeneW, HGNC:5155; HPN.
DR MIM, 142440; -.
DR GO, GO:0005887; C: integral to plasma membrane; TAS.
DR GO, GO:0008151; P: cell growth and/or maintenance; TAS.
DR InterPro, IPR01314; Chymotrypsin.
DR InterPro, IPR01354; Ser. protease_Try.
DR Pfam, PF00089; trypsin.1.
DR PRINTS, PR00722; CHYMOTRYPSIN.
DR SMART, SM00020; TRYPSIN_SPC.1.
DR PROSITE, PS00240; TRYPSIN_DOM.1.
DR PROSITE, PS00134; TRYPSIN_HIS.1.
DR PROSITE, PS00135; TRYPSIN_SER.1.
DR Hydrolase, Serine protease; Transmembrane; Signal-anchor.
FT CHAIN 1 162
FT CHAIN 163 417
FT CHAIN 418 417
FT DOMAIN 1 17
FT TRANSMEM 16 44
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).

```



```

FT DOMAIN 45 417 EXTRACELLULAR (POTENTIAL).
FT 163 417 SERINE PROTEASE.
FT ACT SITE 203 203 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 257 257 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 353 353 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 153 277 INTERCHAIN (BY SIMILARITY).
FT DISULFID 188 204 BY SIMILARITY.
FT DISULFID 322 338 BY SIMILARITY.
FT DISULFID 349 381 BY SIMILARITY.
FT CARBOHYD 112 112 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 417 AA; 45011 MW; B2086F6E1B551D7 CRC64;

Query Match 33.7%; Score 520.5; DB 1; Length 417;
Best Local Similarity 38.7%; Pred. No. 1.4e-39;
Matches 104; Conservative 43; Mismatches 109; Indels 13; Gaps 4;

QY 22 CSSKAVSLRCLTACGNNLSSRSRIVYGESALPGAMPQVSLHVNHVCGSITTPW 81
DB 140 CRRGRFLAICDDCG--RRRLPVDRIVGGRDTSIGWPMQVSLRYDGAHLGGSLASGDW 197
QY 82 IYTAACVCEKPLNPMHTAFAGILNOSFMFGAGQVEKVIHPNY-----DSKTKNN 135
DB 198 VILTAHCFPERNRVLSRMVFAVAQVQASP-HGLQGVQAVVYHGGYLPFRDPNSSENSN 256
QY 136 DIALMKLQKPLTFNDLVKVCLEPNPGMMLQPBOLCISGWLAEKGTSEVINAARVLL 195
DB 257 DIALVHLSSPLPLTERIVQVCLPAAGALVDGKICLVYTGKNTQYVGGQAGVQEARVP 316
QY 196 IETORCNSRYVDNLITPMACAGFLQGVNDSCGDSGGKLV---TSKNKNIWLTGDRS 251
DB 317 ISNDVCGADFGYGNOKPKPMFCAGYEGGIDACGDSGGPFVCEDSISRPRRLGQIYS 376
QY 252 WSGGAKAVRPGVGVGMVFTMIYKMR 280
DB 377 WGTGALAKQKPGVYTKVSDPFRMTFOAIK 405

RESULT 14
FAIL_HUMAN STANDARD; PRT; 625 AA.
AC P03951; Q9Y495;
DT 23-OCT-1986 (Rel. 02, Created)
DT 23-OCT-1986 (Rel. 02, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Coagulation factor XI precursor (EC 3.4.21.27) (Plasma thromboplasmin
DE antecedent) (PFA) (FXI).
GN P11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxId=9606;
RN 11.
RS SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=86243360; PubMed=3636155;
RA Fujikawa K., Chung D.W., Hendrickson L.E., Davie E.W.;
RT "Amino acid sequence of human factor XI, a blood coagulation factor
RT with four tandem repeats that are highly homologous with plasma
RT prekallikrein."
RL Biochemistry 25:2417-2424(1986).
RN 12.
RS SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=88107663; PubMed=2827746;
RA Asakai R., Davie E.W., Chung D.W.;
RT "Organization of the gene for human factor XI."
RL Biochemistry 26:7221-7228(1987).
RN 13.
RS SEQUENCE FROM N.A. (ISOFORM 2).
RX MEDLINE=98256306; PubMed=9593722;
RA Hsu T.-C., Shure S.K., Sesishima T., Bagaera O., Walsh P.N.;
RT "Molecular cloning of platelet factor XI, an alternative splicing
RT product of the plasma factor XI gene."
RL J. Biol. Chem. 273:13787-13793(1998).
[4]

```

```

RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX MEDLINE=91152017; PubMed=1998667;
RA McMullen B.A., Fujikawa K., Davie E.W.;
RT "Location of the disulfide bonds in human coagulation factor XI: the
RT presence of tandem apple domains."
RL Biochemistry 30:2056-2060(1991).
RN 15.
RS VARIANT LEU-301.
RX MEDLINE=90046656; PubMed=2813350;
RA Asakai R., Chung D.W., Ratnoff O.D., Davie E.W.;
RT "Factor XI (Plasma thromboplasmin antecedent) deficiency in Ashkenazi
RT Jews is a bleeding disorder that can result from three types of point
RT mutations."
RL Proc. Natl. Acad. Sci. U.S.A. 86:7667-7671(1989).
RN 16.
RS VARIANT LEU-301.
RX MEDLINE=92190478; PubMed=1547342;
RA Meljers J.C., Davie E.W., Chung D.W.;
RT "Expression of human blood coagulation factor XI: characterization of
RT the defect in factor XI type III deficiency."
RL Blood 79:1435-1440(1992).
RN 17.
RS VARIANTS HIS-34; PRO-320; ILE-322 AND LYS-341.
RX MEDLINE=95195217; PubMed=788672;
RA Pugh R.E., McVey J.H., Tuddenham E.G., Hancock J.F.;
RT "Six point mutations that cause factor XI deficiency."
RL Blood 85:1509-1516(1995).
RN 18.
RS VARIANT VAL-460.
RX MEDLINE=95399291; PubMed=7669672;
RA Imanka Y., Lai K., Nishimura T., Bolton-Maggs P.H., Tuddenham E.G.,
RA McVey J.H.;
RT "Identification of two novel mutations in non-Jewish factor XI
RT deficiency."
RL Br. J. Haematol. 90:916-920(1995).
RN 19.
RS VARIANT ASN-404.
RX MEDLINE=98062240; PubMed=9401068;
RA Wistinghausen B., Reischer A., Oddoux C., Osterer H., Nardi M.,
RA Karpackin M.;
RT "Severe factor XI deficiency in an Arab family associated with a novel
RT mutation in exon 11."
RL Br. J. Haematol. 99:575-577(1997).
RN 110.
RS VARIANTS ARG-244 AND ASN-266.
RX MEDLINE=99005359; PubMed=9787168;
RA Martincic D., Zimmerman S.A., Ware R.E., Sun M.-F., Whitlock J.A.,
RA Gallani D.;
RT "Identification of mutations and polymorphisms in the factor XI genes
RT of an African American family by diideoxyfingerprinting."
RL Blood 92:3309-3317(1998).
RN 111.
RS VARIANT ARG-244.
RX MEDLINE=99150037; PubMed=10027710;
RA Alhag A., Mitchell M., Sethi M., Rahman S., Flynn G., Boulton P.,
RA Caeno G., Smith M., Savidge G.;
RT "Identification of a novel mutation in a non-Jewish factor XI
RT deficient kindred."
RL Br. J. Haematol. 104:44-49(1999).
RN 113.
RS VARIANTS CYS-326; VAL-430 AND ARG-594.
RX MEDLINE=20074632; PubMed=10606881;
RA Mitchell M., Cutler J., Thompson S., Moore G., Jenkins Ap Rees E.,
RA Smith W., Savidge G., Alhag A.;
RT "Heterozygous factor XI deficiency associated with three novel
RT mutations."
RL Br. J. Haematol. 107:763-765(1999).
[14]

```


